

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : <b>G01N 33/574, C12Q 1/68</b>		<b>A2</b>	(11) International Publication Number: <b>WO 98/53319</b> (43) International Publication Date: 26 November 1998 (26.11.98)
(21) International Application Number: <b>PCT/US98/10277</b> (22) International Filing Date: 20 May 1998 (20.05.98)  (30) Priority Data: 60/047,352 21 May 1997 (21.05.97) <b>US</b>  (63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application <b>US</b> 60/047,352 (CON) Filed on 21 May 1997 (21.05.97)  (71) Applicant (for all designated States except US): <b>THE JOHNS HOPKINS UNIVERSITY [US/US]; Suite 2-100, 2024 E. Monument Street, Baltimore, MD 21205 (US).</b>  (72) Inventors; and (75) Inventors/Applicants (for US only): <b>VOGELSTEIN, Bert [US/US]; The Johns Hopkins University, Suite 2-100, 2024 E. Monument Street, Baltimore, MD 21205 (US). KINZLER, Kenneth, W. [US/US]; The Johns Hopkins University, Suite 2-100, 2024 E. Monument Street, Baltimore, MD 21205 (US).</b>		(74) Agents: <b>KAGAN, Sarah, A. et al.; Banner &amp; Witcoff, Ltd., 11th floor, 1001 G Street, N.W., Washington, DC 20001-4597 (US).</b>  (81) Designated States: <b>AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).</b>  <b>Published</b> <i>Without international search report and to be republished upon receipt of that report.</i>	
(54) Title: <b>GENE EXPRESSION PROFILES IN NORMAL AND CANCER CELLS</b>			
(57) Abstract <p>As a step towards understanding the complex differences between normal and cancer cells, gene expression patterns were examined in gastrointestinal tumors. More than 300,000 transcripts derived from at least 45,000 different genes were analyzed. Although extensive similarity was noted between the expression profiles, more than 500 transcripts that were expressed at significantly different levels in normal and neoplastic cells were identified. These data provide insights into the extent of expression differences underlying malignancy and reveal genes that are useful as diagnostic or prognostic markers.</p>			

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Larvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## Gene Expression Profiles in Normal and Cancer Cells

This invention was made with support from the National Institutes of Health, Grant No. GM07309, CA57345, and CA62924. The U.S. government therefore retains certain rights in the invention.

5

### TECHNICAL FIELD OF THE INVENTION

This invention is related to the diagnosis of cancer, and tools for carrying out such diagnosis.

### BACKGROUND OF THE INVENTION

10

Much of cancer research over the past 50 years has been devoted to the analyses of genes that are expressed differently in tumor cells compared to their normal counterparts. Although hundreds of studies have pointed out differences in the expression of one or a few genes, no comprehensive study of gene expression in the cancer cell has been reported. It is therefore not known how many genes are expressed differentially in tumor versus normal cells, whether the bulk of these differences are cell autonomous rather than being dependent on the tumor microenvironment, and whether most differences are cell-type specific or tumor specific. Thus there is a need in the art for information on the molecular changes that occur in cells during cancer development and progression.

15

**SUMMARY OF THE INVENTION**

According to one embodiment of the invention, a method is provided for diagnosing colon cancer in a sample suspected of being neoplastic. The method comprises the steps of:

5                   comparing the level of at least one transcript in a first sample of a tissue to a second sample, wherein the first sample is of a colonic tissue suspected of being neoplastic and the second sample is of a normal human colonic tissue, and wherein the transcript is identified by a tag selected from the group consisting of those shown in Table 3;

10                   identifying the first sample as neoplastic when the level of the at least one transcript is found to be lower in the first sample than in the second sample.

According to another embodiment of the invention, another method is provided for diagnosing colon cancer in a sample suspected of being neoplastic. The method comprises the steps of:

15                   comparing the level of at least one transcript in a first sample of a tissue to a second sample, wherein the first sample is of a colonic tissue suspected of being neoplastic and the second sample is of a normal human colonic tissue, and wherein the transcript is identified by a tag selected from the group consisting of those shown in Table 2;

20                   identifying the first sample as neoplastic when the level of the at least one transcript is found to be higher in the first sample than in the second sample.

In another embodiment of the invention an isolated and purified human nucleic acid molecule is provided. The molecule comprises a SAGE tag selected from SEQ ID NO:1-732.

25                   In yet another aspect of the invention an isolated nucleotide probe is provided. The probe comprises at least 12 nucleotides of a human nucleic acid molecule, wherein the human nucleic acid molecule comprises a SAGE tag selected from SEQ ID NO: 1-732.

According to another aspect of the invention a method is provided for diagnosing pancreatic cancer in a sample suspected of being neoplastic. The method comprises the steps of:

5       comparing the level of at least one transcript in a first sample of a tissue to a second sample, wherein the first sample is of a pancreatic tissue suspected of being neoplastic and the second sample is of a normal human colon tissue, wherein said transcript is identified by a tag selected from the group consisting of those shown Table 4;

10       identifying the first sample as neoplastic when the level of the at least one transcript is found to be higher in the first sample than in the second sample.

According to still another embodiment of the invention a method of diagnosing cancer in a sample suspected of being neoplastic is provided. The method comprises the steps of:

15       comparing the level of at least one transcript in a first sample of a tissue to a second sample, wherein the first sample is of a tissue suspected of being neoplastic and the second sample is of a normal human tissue, wherein said transcript is identified by a tag selected from the group consisting of those shown Table 5;

20       identifying the first sample as neoplastic when the level of the at least one transcript is found to be higher in the first sample than in the second sample.

According to another embodiment of the invention a method is provided to aid in the determination of a prognosis for a colon cancer patient.

25       The method comprises the steps of:

30       comparing the level of at least one transcript in a first sample of a tissue to a second sample, wherein the first sample is of a neoplastic colonic tissue and the second sample is of a normal human colonic tissue, and wherein the transcript is identified by a tag selected from the group consisting of those shown in Table 3;

determining a poorer prognosis if the level of the at least one transcript is found to be lower in the first sample than in the second sample.

According to another aspect of the invention a method to aid in determining a prognosis for a patient with colon cancer is provided. The method comprises the steps of:

comparing the level of at least one transcript in a first tissue sample to a second sample, wherein the first sample is of a colonic cancer tissue and the second sample is of a normal human colonic tissue, and wherein the transcript is identified by a tag selected from the group consisting of those shown in Table 2;

determining a poorer prognosis if the level of the at least one transcript is found to be higher in the first sample than in the second sample.

In yet another embodiment of the invention a method is provided for diagnosing colon cancer in a sample suspected of being neoplastic. The method comprises the steps of:

comparing the level of expression of at least one protein in a first sample of a tissue to a second sample, wherein the first sample is of a colonic tissue suspected of being neoplastic and the second sample is of a normal human colonic tissue, and wherein the protein is encoded by a transcript identified by a tag selected from the group consisting of those shown in Table 3;

identifying the first sample as neoplastic when the level of expression of the protein is found to be lower in the first sample than in the second sample.

In another aspect of the invention a method of diagnosing colon cancer in a sample suspected of being neoplastic is provided. The method comprises the steps of:

comparing the level of expression of at least one protein in a first sample of a tissue to a second sample, wherein the first sample is of a colonic tissue suspected of being neoplastic and the second sample is of a normal human colonic tissue, and wherein the protein is encoded by a transcript

identified by a tag selected from the group consisting of those shown in Table 2;

identifying the first sample as neoplastic when expression of the protein is found to be higher in the first sample than in the second sample.

5 According to another embodiment of the invention a method is provided to aid in determining a prognosis of a patient having pancreatic cancer. The method comprises the steps of:

10 comparing the level of at least one transcript in a first sample of a tissue to a second sample, wherein the first sample is of a neoplastic pancreatic tissue and the second sample is of a normal human colon tissue, wherein said transcript is identified by a tag selected from the group consisting of those shown Table 4;

determining a poorer prognosis if transcription is found to be higher in the first sample than in the second sample.

15 In yet another aspect of the invention a method to aid in providing a prognosis for a cancer patient is provided. The method comprises the steps of:

20 comparing the level of at least one transcript in a first sample of a tissue to a second sample, wherein the first sample is of a neoplastic tissue and the second sample is of a normal human tissue of the same tissue type, wherein said transcript is identified by a tag selected from the group consisting of those shown Table 5;

determining a poorer prognosis if transcription is found to be higher in the first sample than in the second sample.

25 According to still another aspect of the invention, a method is provided for diagnosing pancreatic cancer in a sample suspected of being neoplastic. The method comprises the steps of:

30 comparing the level of expression of at least one protein encoded by a transcript in a first sample of a tissue to a second sample, wherein the first sample is of a pancreatic tissue suspected of being neoplastic and the second sample is of a normal human colon tissue, wherein said protein is

encoded by a transcript identified by a tag selected from the group consisting of those shown Table 4;

identifying the first sample as neoplastic when expression of the protein is found to be higher in the first sample than in the second sample.

5 According to yet another aspect of the invention a method is provided for diagnosing cancer in a sample suspected of being neoplastic. The method comprises the steps of:

10 comparing the level of expression of at least one protein in a first sample of a tissue to a second sample, wherein the first sample is of a tissue suspected of being neoplastic and the second sample is of a normal human tissue, wherein said protein is encoded by a transcript identified by a tag selected from the group consisting of those shown Table 5;

identifying the first sample as neoplastic when expression of the protein is found to be higher in the first sample than in the second sample.

15 In still another embodiment of the invention a method is provided to aid in the determination of a prognosis of a colon cancer patient. The method comprises the steps of:

20 comparing the level of expression of at least one protein in a first sample of a tissue to a second sample, wherein the first sample is of a neoplastic colonic tissue and the second sample is of a normal human colonic tissue, and wherein the protein is encoded by a transcript identified by a tag selected from the group consisting of those shown in Table 3;

determining a poorer prognosis if the level of expression is found to be lower in the first sample than in the second sample.

25 In still another embodiment of the invention a method is provided to aid in determining a prognosis for a patient with colon cancer. The method comprises the steps of:

30 comparing the level of expression of at least one protein in a first tissue sample to a second sample, wherein the first sample is of a colonic cancer tissue and the second sample is of a normal human colonic tissue, and



wherein the protein is encoded by a transcript identified by a tag selected from the group consisting of those shown in Table 2;

determining a poorer prognosis if the level of expression is found to be higher in the first sample than in the second sample.

5 In still another aspect of the invention a method is provided to aid in determining a prognosis of a patient having pancreatic cancer. The method comprises the steps of:

10 comparing the level of expression of at least one protein in a first sample of a tissue to a second sample, wherein the first sample is of a neoplastic pancreatic tissue and the second sample is of a normal human colon tissue, wherein said protein is encoded by a transcript identified by a tag selected from the group consisting of those shown Table 4;

determining a poorer prognosis if the level of expression is found to be higher in the first sample than in the second sample.

15 According to even a further aspect of the invention a method is provided to aid in providing a prognosis for a cancer patient. The method comprises the steps of:

20 comparing the level of expression of at least one protein in a first sample of a tissue to a second sample, wherein the first sample is of a neoplastic tissue and the second sample is of a normal human tissue of the same tissue type, wherein said protein is encoded by a transcript identified by a tag selected from the group consisting of those shown Table 5;

determining a poorer prognosis if the level of expression is found to be higher in the first sample than in the second sample.

25 In still another embodiment of the invention a method of treating a cancer cell is provided. The method comprises the step of:

30 administering to a cancer cell an antibody which specifically binds to a protein encoded by a transcript identified by a tag selected from the group consisting of those shown in Tables 2, 4, and 5, wherein the antibody is linked to a cytotoxic agent.

In another aspect of the invention an antibody linked to a cytotoxic agent is provided. The antibody specifically binds to a protein encoded by a transcript identified by a tag selected from the group consisting of those shown in Tables 2, 4, and 5.

5 According to another aspect of the invention, a method of detecting colon cancer in a patient is provided. The method comprises the steps of:

10 comparing the level of at least one protein or transcript in a first body sample to a second body sample, wherein the first sample is a body sample of the patient and the second sample is of a normal human, wherein the protein is encoded by a transcript and the transcript is identified by a tag selected from the group consisting of those shown in Table 2, wherein the first and second body sample is a sample selected from the group consisting of blood, urine, feces, sputum, and serum;

15 identifying neoplasia when the level of the at least one protein or transcript is found to be higher in the first sample than in the second sample.

In another aspect of the invention a method of detecting pancreatic cancer in a patient is provided. The method comprises the steps of:

20 comparing the level of at least one protein or transcript encoded by a transcript in a first sample of a tissue to a second sample, wherein the first sample is of the patient and the second sample is of a normal human, wherein said protein is encoded by a transcript and the transcript is identified by a tag selected from the group consisting of those shown Table 4, wherein the first and second sample is a sample selected from the group consisting of blood, urine, feces, sputum, and serum;

25 identifying neoplasia when the level of the at least one protein or transcript is found to be higher in the first sample than in the second sample.

Also provided by the present invention is a method of detecting cancer in a patient. The method comprises the steps of:

30 comparing the level of at least one protein or transcript in a first sample to a second sample, wherein the first sample is of patient and the second sample is of a normal human, wherein said protein is encoded by a

transcript and the transcript is identified by a tag selected from the group consisting of those shown Table 5, wherein the first and second body sample is a sample selected from the group consisting of blood, urine, feces, sputum, and serum;

5 identifying neoplasia when the level of the at least one protein or transcript is found to be higher in the first sample than in the second sample.

Additionally provided by the present invention is a method to aid in the determination of a prognosis for a colon cancer patient. The method comprises the steps of:

10 comparing the level of at least one protein or transcript in a first sample to a second sample, wherein the first sample is of a colon cancer patient and the second sample is of a normal human, wherein the protein is encoded by a transcript and the transcript is identified by a tag selected from the group consisting of those shown in Table 3, wherein the first and second body sample is a sample selected from the group consisting of blood, urine, feces, sputum, and serum;

15 determining a poorer prognosis if the level of the at least one protein or transcript is found to be lower in the first sample than in the second sample.

20 Provided by another embodiment of the invention is a method to aid in determining a prognosis for a patient with colon cancer. The method comprises the steps of:

25 comparing the level of at least one protein or transcript in a first sample to a second sample, wherein the first sample is of a colonic cancer patient and the second sample is of a normal human, wherein the protein is encoded by a transcript and the transcript is identified by a tag selected from the group consisting of those shown in Table 2, wherein the first and second sample is a sample selected from the group consisting of blood, urine, feces, sputum, and serum;

determining a poorer prognosis if the level of the at least one protein or transcript is found to be higher in the first sample than in the second sample.

5 According to still another aspect of the invention, a method to aid in determining a prognosis of a patient having pancreatic cancer is provided. The method comprises the steps of:

10 comparing the level of at least one protein or transcript in a first sample to a second sample, wherein the first sample is of a pancreatic cancer patient and the second sample is of a normal human, wherein said protein is encoded by a transcript and the transcript is identified by a tag selected from the group consisting of those shown Table 4, wherein said first and second sample is a sample selected from the group consisting of blood, urine, feces, sputum, and serum;

15 determining a poorer prognosis if the level of the at least one protein or transcript is found to be higher in the first sample than in the second sample.

Also provided by the present invention is a method to aid in providing a prognosis for a cancer patient. The method comprises the steps of:

20 comparing the level of expression of at least one protein or transcript in a first sample to a second sample, wherein the first sample is of a cancer patient and the second sample is of a normal human, wherein said protein is encoded by a transcript and the transcript is identified by a tag selected from the group consisting of those shown Table 5, wherein the first and second sample is a sample selected from the group consisting of blood, 25 urine, feces, sputum, and serum;

determining a poorer prognosis if the level of the at least one protein or transcript is found to be higher in the first sample than in the second sample.

30 The present invention further includes antisense oligonucleotides complementary in whole or in part to SEQ ID NOS:1-732.

This invention also provides a method for screening for candidate agents that modulate the expression of a polynucleotide selected from the group consisting of the polynucleotides in SEQ ID NOS.1-732 or their respective complements, by contacting a test agent with a pancreatic or colon cell and monitoring expression of the polynucleotide, wherein the test agent which modifies the expression of the polynucleotide is a candidate agent.

The present invention provides the art with new methods and reagents for diagnosing and prognosing cancers. In addition, some of the newly disclosed genes may play an important role in the development of cancers.

### **BRIEF DESCRIPTION OF THE DRAWINGS**

**Fig. 1.** Comparison of expression patterns in colorectal cancers and normal colon epithelium. **(FIG. 1A)** A semi-logarithmic plot reveals 51 tags that were decreased more than 10 fold in primary CR cancer cells whereas 32 tags were increased more than 10 fold. 62,168 and 60,878 tags derived from normal colon epithelium and primary CR cancers, respectively, were used for this analysis. The relative expression of each transcript was determined by dividing the number of tags observed in tumor and normal tissue as indicated. To avoid division by 0, a tag value of 1 was used for any tag that was not detectable in one of the samples. These ratios were then rounded to the nearest integer and their distribution plotted on the abscissa. The number of genes displaying each ratio was plotted on the ordinate. Tu: CR tumors; NC: Normal colon. **(FIG. 1B and FIG. 1C)** Differentially expressed genes in colorectal cancers. The number of transcripts found to be differentially expressed ( $P < 0.01$ ) are presented as Venn diagrams. Diagrams of transcripts that were decreased **(FIG. 1B)** or increased **(FIG. 1C)** in CR cancers compared to normal colon epithelium. Comparisons were between primary tumors and cells in culture as indicated.

**Fig. 2.** Northern blot analysis of genes differentially expressed in gastrointestinal neoplasia. Northern blot analysis was performed on total RNA (5  $\mu$ g isolated from primary CR carcinomas (T) and matching normal colon epithelium (N), or pancreatic carcinomas. The top panel in each case show an

example of the ethidium bromide stained gels prior to transfer. The number of SAGE tags observed in the original analysis is indicated to the right of each blot. (FIG. 2A) Examples of transcripts that were decreased or increased in CR cancers. (FIG. 2B) Examples of transcripts increased in pancreatic cancers (10). (FIG. 2C) Examples of transcripts elevated in cancer which were or were not cancer type specific. Probes used for Northern blot analysis were as follows (Human SAGE Tag unique identifier, gene name, (GenBank accession number)): (FIG. 2A) H204104, Guanylin (M95714); H259108, (see Table 2); H1000193, (see Table 2); H998030, (see Table 2). (FIG. 2B) H294155, RIG-E (U42376); H560056, TIMP-1 (S68252). (FIG. 2C) H802810, EST338411 (W52120); H85882, 1-8D (X57351); H618841, GA733-1 (X13425).

**Tables 2-5. Transcripts Differentially Expressed in Human Cancer.**

Tag sequence represents the NlaIII site plus the adjacent 11 bp SAGE tag. Tag number indicates a SAGE UID (unique identifier). NC, TU, CL, PT, PC, refers to the number of the indicated tag observed in RNA isolated from normal colorectal epithelium, primary colorectal cancers, colorectal cancer cell lines, primary pancreatic cancers, or pancreatic cancer cell lines, respectively. The Accession and Gene Name refer to representative GenBank entries that contain the tag sequence.

Table 2 Transcripts increased in colorectal cancer.

Table 3 Transcripts decreased in colorectal cancer.

Table 4 Transcripts increased in pancreatic cancer.

Table 5 Transcripts increased in pancreatic and colorectal cancer.

**DETAILED DESCRIPTION**

The inventors have discovered sets of human genes which are either upregulated or downregulated in cancer cells, as compared to normal cells. Specifically, certain genes have been found to be upregulated or downregulated in colorectal and/or pancreatic cancer cells, when compared to normal colon

cells. These sets of differentially regulated genes can be used as diagnostic markers, either individually or in sets of, for example, 2, 5, 10, 20, or 30.

Genes whose expression was detected to be increased in colorectal cancer are shown in Table 2. Genes whose expression was detected to be decreased in colorectal cancer are shown in Table 3. Genes whose expression was detected as increased in pancreatic cancer are shown in Table 4. Genes whose expression was detected as increased in both pancreatic cancer and colorectal cancer are shown in Table 5. These latter genes likely play a role in neoplastic development generally.

Tag sequences, as provided herein, uniquely identify genes. This is due to their length, and their specific location (3') in a gene from which they are drawn. The full length genes can be identified by matching the tag to a gene data base member, or by using the tag sequences as probes to physically isolate previously unidentified genes from cDNA libraries. The methods by which genes are isolated from libraries using DNA probes are well known in the art. See, for example, Veculescu et al., *Science* 270: 484 (1995), and Sambrook et al. (1989), *MOLECULAR CLONING: A LABORATORY MANUAL*, 2nd ed. (Cold Spring Harbor Press, Cold Spring Harbor, New York). Once a gene or transcript has been identified, either by matching to a data base entry, or by physically hybridizing to a cDNA molecule, the position of the hybridizing or matching region in the transcript can be determined. If the tag sequence is not in the 3' end, immediately adjacent to the restriction enzyme used to generate the SAGE tags, then a spurious match may have been made. Confirmation of the identity of a SAGE tag can be made by comparing transcription levels of the tag to that of the identified gene in certain cell types.

In addition to the sequences shown in SEQ ID NOS: 1-732, or their complements, this invention also provides the anti-sense polynucleotide stand, e.g. antisense RNA to these sequences or their complements. One can obtain an antisense RNA using the sequences provided in SEQ ID NOS: 1-732 and the methodology described in Vander Krol et al. (1988) *BioTechniques* 6:958.

The invention also encompasses polynucleotides which differ from that of the polynucleotides described above, but which produce the same phenotypic effect, such as the allele. These altered, but phenotypically equivalent polynucleotides are referred to "equivalent nucleic acids." This invention also encompasses polynucleotides characterized by changes in non-coding regions that do not alter the phenotype of the polypeptide produced therefrom when compared to the polynucleotide herein. This invention further encompasses polynucleotides, which hybridize to the polynucleotides of the subject invention under conditions of moderate or high stringency.

The polynucleotides can be conjugated to a detectable marker, e.g., an enzymatic label or a radioisotope for detection of nucleic acid and/or expression of the gene in a cell. A wide variety of appropriate detectable markers are known in the art, including fluorescent, radioactive, enzymatic or other ligands, such as avidin/biotin, which are capable of giving a detectable signal. In preferred embodiments, one will likely desire to employ a fluorescent label or an enzyme tag, such as urease, alkaline phosphatase or peroxidase, instead of radioactive or other environmental undesirable reagents. In the case of enzyme tags, colorimetric indicator substrates are known which can be employed to provide a means visible to the human eye or spectrophotometrically, to identify specific hybridization with complementary nucleic acid-containing samples. Briefly, this invention further provides a method for detecting a single-stranded polynucleotide identified by SEQ ID NOS.1-732 or its complement, by contacting target single-stranded polynucleotides with a labeled, single-stranded polynucleotide (a probe) which is at least 10 nucleotides of the complement of SEQ ID NOS: 1-732 (or the corresponding complement) under conditions permitting hybridization (preferably moderately stringent hybridization conditions) of complementary single-stranded polynucleotides, or more preferably, under highly stringent hybridization conditions. Hybridized polynucleotide pairs are separated from un-hybridized, single-stranded polynucleotides. The hybridized polynucleotide



pairs are detected using methods well known to those of skill in the art and set forth, for example, in Sambrook et al. (1989) *supra*.

5 The polynucleotides of this invention can be isolated using the technique described in the experimental section or replicated using PCR. The PCR technology is the subject matter of United States Patent Nos. 4,683,195, 4,800,159, 4,754,065, and 4,683,202 and described in PCR: The Polymerase Chain Reaction (Mullis et al. eds, Birkhauser Press, Boston (1994)) or MacPherson et al. (1991) and (1994), *supra*, and references cited therein. Alternatively, one of skill in the art can use the sequences provided herein and a commercial DNA synthesizer to replicate the DNA. Accordingly, this invention also provides a process for obtaining the polynucleotides of this invention by providing the linear sequence of the polynucleotide, nucleotides, appropriate primer molecules, chemicals such as enzymes and instructions for their replication and chemically replicating or linking the nucleotides in the proper orientation to obtain the polynucleotides. In a separate embodiment, these polynucleotides are further isolated. Still further, one of skill in the art can insert the polynucleotide into a suitable replication vector and insert the vector into a suitable host cell (procaryotic or eucaryotic) for replication and amplification. The DNA so amplified can be isolated from the cell by methods well known to those of skill in the art. A process for obtaining polynucleotides by this method is further provided herein as well as the polynucleotides so obtained.

25 RNA can be obtained by first inserting a DNA polynucleotide into a suitable host cell. The DNA can be inserted by any appropriate method, e.g., by the use of an appropriate gene delivery vector or by electroporation. When the cell replicates and the DNA is transcribed into RNA; the RNA can then be isolated using methods well known to those of skill in the art, for example, as set forth in Sambrook et al. (1989) *supra*. For instance, mRNA can be isolated using various lytic enzymes or chemical solutions according to the procedures set forth in Sambrook et al. (1989), *supra* or extracted by nucleic-acid-binding resins following the accompanying instructions provided by manufactures.

Polynucleotides having at least 10 nucleotides and exhibiting sequence complementarity or homology to SEQ ID NOS: 1-732 find utility as hybridization probes. In some aspects, the full coding sequence of the transcript, i.e., for SEQ ID NOS: 1-732, are known. Accordingly, any portion of the known sequences available in GenBank, or homologous sequences, can be used in the methods of this invention.

It is known in the art that a "perfectly matched" probe is not needed for a specific hybridization. Minor changes in probe sequence achieved by substitution, deletion or insertion of a small number of bases do not affect the hybridization specificity. In general, as much as 20% base-pair mismatch (when optimally aligned) can be tolerated. Preferably, a probe useful for detecting the aforementioned mRNA is at least about 80% identical to the homologous region of comparable size contained in the previously identified sequences identified by SEQ ID NOS:1-732, which correspond to previously characterized genes or SEQ ID NOS:1-732, which correspond to known ESTs. More preferably, the probe is 85% identical to the corresponding gene sequence after alignment of the homologous region; even more preferably, it exhibits 90% identity.

These probes can be used in radioassays (e.g. Southern and Northern blot analysis) to detect, prognose, diagnose or monitor various pancreatic or colon cells or tissue containing these cells. The probes also can be attached to a solid support or an array such as a chip for use in high throughput screening assays for the detection of expression of the gene corresponding to one or more polynucleotide(s) of this invention. Accordingly, this invention also provides at least one of the transcripts identified as SEQ ID NOS:1-732, or its complement, attached to a solid support for use in high throughput screens.

The total size of fragment, as well as the size of the complementary stretches, will depend on the intended use or application of the particular nucleic acid segment. Smaller fragments will generally find use in hybridization embodiments, wherein the length of the complementary region may be varied,

such as between about 10 and about 100 nucleotides, or even full length according to the complementary sequences one wishes to detect.

5 Nucleotide probes having complementary sequences over stretches greater than 10 nucleotides in length are generally preferred, so as to increase stability and selectivity of the hybrid, and thereby improving the specificity of particular hybrid molecules obtained. More preferably, one can design polynucleotides having gene-complementary stretches of more than 50 nucleotides in length, or even longer where desired. Such fragments may be readily prepared by, for example, directly synthesizing the fragment by  
10 chemical means, by application of nucleic acid reproduction technology, such as the PCR technology with two priming oligonucleotides as described in U.S. Pat. No. 4,603,102 or by introducing selected sequences into recombinant vectors for recombinant production. A preferred probe is about 50-75 or more preferably, 50-100, nucleotides in length.

15 The polynucleotides of the present invention can serve as primers for the detection of genes or gene transcripts that are expressed in pancreatic or colon cells. In this context, amplification means any method employing a primer-dependent polymerase capable of replicating a target sequence with reasonable fidelity. Amplification may be carried out by natural or recombinant DNA-polymerases such as T7 DNA polymerase, Klenow fragment of E.coli  
20 DNA polymerase, and reverse transcriptase.

A preferred amplification method is PCR. However, PCR conditions used for each reaction are empirically determined. A number of parameters influence the success of a reaction. Among them are annealing temperature and time, extension time,  $Mg^{2+}$  ATP concentration, pH, and the relative  
25 concentration of primers, templates, and deoxyribonucleotides. After amplification, the resulting DNA fragments can be detected by agarose gel electrophoresis followed by visualization with ethidium bromide staining and ultraviolet illumination.

30 The invention further provides the isolated polynucleotide operatively linked to a promoter of RNA transcription, as well as other regulatory

sequences for replication and/or transient or stable expression of the DNA or RNA. As used herein, the term "operatively linked" means positioned in such a manner that the promoter will direct transcription of RNA off the DNA molecule. Examples of such promoters are SP6, T4 and T7. In certain  
5       embodiments, cell-specific promoters are used for cell-specific expression of the inserted polynucleotide. Vectors which contain a promoter or a promoter/enhancer, with termination codons and selectable marker sequences, as well as a cloning site into which an inserted piece of DNA can be operatively linked to that promoter are well known in the art and commercially available.  
10       For general methodology and cloning strategies, see Gene Expression Technology (Goeddel ed., Academic Press, Inc. (1991)) and references cited therein and Vectors: Essential Data Series (Gacsa and Ramji, eds., John Wiley & Sons, N.Y. (1994)), which contains maps, functional properties, commercial suppliers and a reference to GenEMBL accession numbers for various suitable  
15       vectors. Preferable, these vectors are capable of transcribing RNA in vitro or in vivo.

Fragment of the sequences shown in SEQ ID NOS:1-732 or their respective complements also are encompassed by this invention, preferably at least 10 nucleotides and more preferably having at least 18 nucleotides. Larger  
20       polynucleotides, e.g., cDNA or genomic DNA, which hybridize under moderate or stringent conditions to the polynucleotide sequences shown in SEQ ID NOS:1-732, or their respective complements, also are encompassed by this invention.

In one embodiment, these fragments are polynucleotides that encode  
25       polypeptides or proteins having diagnostic and therapeutic utilities as described herein as well as probes to identify transcripts of the protein which may or may not be present. These nucleic acid fragments can be prepared, for example, by restriction enzyme digestion of the polynucleotide of SEQ ID NOS:1-732, or their complements, and then labeled with a detectable marker. Alternatively,  
30       random fragments can be generated using nick translation of the molecule. For

methodology for the preparation and labeling of such fragments, see Sambrook et al., (1989) supra.

5 Expression vectors containing these nucleic acids are useful to obtain host vector systems to produce proteins and polypeptides. It is implied that these expression vectors must be replicable in the host organisms either as episomes or as an integral part of the chromosomal DNA. Suitable expression  
10 vectors include viral vectors, including adenoviruses, adeno-associated viruses, retroviruses, cosmids, etc. Adenoviral vectors are particularly useful for introducing genes into tissues in vivo because of their high levels of expression and efficient transformation of cells both in vitro and in vivo. When a nucleic  
15 acid is inserted into a suitable host cell, e.g., a procaryotic or a eucaryotic cell and the host cell replicates, the protein can be recombinantly produced. Suitable host cells will depend on the vector and can include mammalian cells, animal cells, human cells, simian cells, insect cells, yeast cells, and bacterial  
20 cells constructed using well known methods. See Sambrook et al. (1989) supra. In addition to the use of viral vector for insertion of exogenous nucleic acid into cells, the nucleic acid can be inserted into the host cell by methods well known in the art such as transformation for bacterial cells; transfection  
25 using calcium phosphate precipitation for mammalian cells; or DEAE-dextran; electroporation; or microinjection. See Sambrook et al. (1989) supra for this methodology. Thus, this invention also provides a host cell, e.g. a mammalian  
30 cell, an animal cell (rat or mouse), a human cell, or a procaryotic cell such as a bacterial cell, containing a polynucleotide encoding a protein or polypeptide or antibody.

25 When the vectors are used for gene therapy *in vivo* or *ex vivo*, a pharmaceutically acceptable vector is preferred, such as a replication-incompetent retroviral or adenoviral vector. Pharmaceutically acceptable vectors containing the nucleic acids of this invention can be further  
30 modified for transient or stable expression of the inserted polynucleotide. As used herein, the term "pharmaceutically acceptable vector" includes, but is not limited to, a vector or delivery vehicle having the ability to selectively target

and introduce the nucleic acid into dividing cells. An example of such a vector is a "replication-incompetent" vector defined by its inability to produce viral proteins, precluding spread of the vector in the infected host cell. An example of a replication-incompetent retroviral vector is LNL6 (Miller, A.D. et al. (1989) BioTechniques 7:980-990). The methodology of using replication-incompetent retroviruses for retroviral-mediated gene transfer of gene markers is well established (Correll et al. (1989) PNAS USA 86:8912; Bordignon (1989) PNAS USA 86:8912-52; Culver, K. (1991) PNAS USA 88:3155; and Rill, D.R. (1991) Blood 79(10):2694-700. Clinical investigations have shown that there are few or no adverse effects associated with the viral vectors, see Anderson (1992) Science 256:808-13.

Compositions containing the polynucleotides of this invention, in isolated form or contained within a vector or host cell are further provided herein. When these compositions are to be used pharmaceutically, they are combined with a pharmaceutically acceptable carrier.

This invention further encompasses genes, either genomic or cDNA, which code for a polypeptide or protein in the cell of interest. The genes specifically hybridize under moderate or stringent conditions to a polynucleotide identified by SEQ ID NOS: 1-732 or their respective complements. The process of identification of larger fragment or the full-length coding sequence to which the partial sequence depicted in SEQ ID NOS:1-732 hybridizes preferably involves the use of the methods and reagents provided in this invention, either singularly or in combination.

Five methods are disclosed herein which allows one of skill in the art to isolate the gene or cDNA corresponding to the transcripts of the invention.

#### RACE-PCR Technique

One method to isolate the gene or cDNA which code for a polypeptide or protein and which corresponds to a transcript of this invention, involves the 5'-RACE-PCR technique. In this technique, the poly-A mRNA that contains the coding sequence of particular interest is first identified by hybridization to

a sequence disclosed herein and then reverse transcribed with a 3'-primer comprising the sequence disclosed herein. The newly synthesized cDNA strand is then tagged with an anchor primer of a known sequence, which preferably contains a convenient cloning restriction site attached at the 5' end. The tagged cDNA is then amplified with the 3'-primer (or a nested primer sharing sequence homology to the internal sequences of the coding region) and the 5'-anchor primer. The amplification may be conducted under conditions of various levels of stringency to optimize the amplification specificity. 5'-RACE-PCR can be readily performed using commercial kits (available from, e.g., BRL Life Technologies Inc, Clontech) according to the manufacturer's instructions.

#### Identification of known genes or ESTs

In addition, databases exist that reduce the complexity of ESTs by assembling contiguous EST sequences into tentative genes. For example, TIGR has assembled human ESTs into a database called THC for tentative human consensus sequences. The THC database allows for a more definitive assignment compared to ESTs alone. Software programs exist (give examples) that allow for assembling ESTs into contiguous sequences from any organism.

#### Isolation of cDNAs from a library by probing with the SAGE transcript or tag

Alternatively, mRNA from a sample preparation was used to construct cDNA library in the ZAP Express vector following the procedure described in Velculescu et al. (1997) Science 270:484. The ZAP Express cDNA synthesis kit (Stratagene) was used accordingly to the manufacturer's protocol. Plates containing 250 to 2000 plaques are hybridized as described in Rupert et al. (1988) Mol. Cell. Bio. 8:3104 to oligonucleotide probes with the same conditions previously described for standard probes except that the hybridization temperature is reduced to room temperature. Washes are performed in 6X standard-saline-citrate 0.1% SDS for 30 minutes at room temperature. The probes are labeled with 32P-ATP through use of T4 polynucleotide kinase.

Table 2 - Transcripts increased in colon cancer

### Transcripts increased in only colon primary tumors compared to normal colon (61 genes)

NC: Normal Colon

TU: Colon Primary Tumor

CL: Colon Cancer Cell Line

PT: Pancreatic Primary Tumor

PC: Pancreatic Cancer Cell Line

#	Tag Sequence	Tag Number	NC	TU	CL	PT	PC	Accession	Gene Name
1	CATGCACCTAATTGG	H285759	612	755	411	161	333	F15516	H.sapiens mitochondrial EST sequence (1-412) from
2	CATGTGATTCACTT	H933704	452	595	235	80	314	U35430	Human cytochrome c oxidase subunit III (COIII) pse
3	CATGCCTGTAA TGCC	H388150	433	549	380	443	197	Z70701	H.sapiens mRNA (fetal brain cDNA c2_11).
								X71347	H.sapiens HNFI-C mRNA.
								X71346	H.sapiens HNFI-B mRNA.
4	CATGCACCTACTCACC	H291282	293	527	78	14	83	U09500	Human mitochondrial cytochrome b gene, partial cds
5	CATGGTGAAACCCCA(G)	H753750	392	517	389	453	194	X66785	H.sapiens mRNA for transacylase (DBT).
								X17648	Human mRNA for granulocyte-macrophage colony-stimu
								U09087	Human thymopoietin beta mRNA, complete cds.
								U09088	Human thymopoietin gamma mRNA, complete cds.
								U20770	Human metastasis suppressor (KAI1) mRNA, complete
								W15552	zb91h11.s1 Soares parathyroid tumor Nbl1PA Homo sap
6	CA1GGGCTTTAGGGA	H687915	37	372	6	29	11	W32091	zc05d03.s1 Soares parathyroid tumor Nbl1PA Homo sap
								R62866	yl11d07.r1 Homo sapiens cDNA clone 138925 5'.
								X89839	H.sapiens mitochondrial DNA for loop attachment se
7	CA1GACTTTTCCAAA	H130369	32	272	32	23	20	T11555	A1486F Homo sapiens cDNA clone A1486 similar to Mi
8	CATGTGGTGTATGCA	H965434	53	271	6	30	5	T15773	IB1870 Homo sapiens cDNA 3'end similar to Human mi
9	CATGAGGGTGTITTC	H175872	26	218	7	20	10	X12544	Human mRNA for HLA class II DR-beta (HLA-DR B).
10	CATGAGGTCAGGAGA(T)	H177315	93	213	113	148	58	S73483	phosphorylase kinase catalytic subunit PHKG2 homol
								X74301	H.sapiens mRNA for MHC class II transactivator.
11	CATGTTGGCCAGGCT	H1025322	124	194	63	111	51	U28687	Human zinc finger containing protein ZNF157 (ZNF15
								U29119	Human leiomyoma LM-196.4 ectopic sequence from HMG
								U36236	Human Fc alpha receptor b mRNA, complete cds.
								W03751	za62h11.r1 Soares fetal liver spleen INFLS Homo sa
12	CATGATCAGCCCTC	H214616	97	186	17	41	49	W03770	za63f10.r1 Soares fetal liver spleen INFLS Homo sa



[illegible]

36	CATGGTGAAACCCA	H753749	9	31	22	30	4	T95857	ye42f01.s1 Homo sapiens cDNA clone 120409 3' simil
								W03237	za35b09.r1 Soares fetal liver spleen INFLS Homo sa
								W03326	za63g03.r1 Soares fetal liver spleen INFLS Homo sa
37	CATGGAAACTGAACA	H526210	6	26	17	5	3	X54195	Human line-1 element DNA, host sequence flanking 1
								U29607	Human methionine aminopeptidase mRNA, complete cds
								H95100	yw57b10.r1 Homo sapiens cDNA clone 256315 5' simil
38	CATGACTTTTTAAAA	H131009	1	22	4	1	0		Human keratinocyte cDNA, clone 067.
39	CATGGAAGTGGTCC	H555450	0	21	7	9	12	D29062	Human keratinocyte cDNA, clone 713.
								D29563	FB3B5 Homo sapiens cDNA clone FB3B5 3'end.
40	CATGTCAGTGGTAGT	H863923	4	21	2	2	1	T03196	H.sapiens CpG DNA, clone 164a10, reverse read cpg1
41	CATGAAACTGTGTT	H7916	2	20	2	2	1	Z57093	H.sapiens CpG island DNA genomic MseI fragment, cl
								Z60184	H.sapiens CpG island DNA genomic MseI fragment, cl
								Z63649	H.sapiens CpG island DNA genomic MseI fragment, cl
								W31349	zb95d06.s1 Soares parathyroid tumor NbHPA Homo sap
42	CATGGGGGGGGGGT	H699051	0	19	0	0	0		zb96i01.s1 Soares parathyroid tumor NbHPA Homo sap
43	CATGGTGCCCGTGCC		2	19	1	0	0	W31448	zb40b06.r1 Soares senescent fibroblasts NbISF Homo
								W47282	H.sapiens fus mRNA.
44	CATGGGGGGTAACTA	H699144	3	19	15	12	5	X71428	TL5=translocated in liposarcoma (human, mRNA, 1824
								S62140	zb96a06.r1 Soares parathyroid tumor NbHPA Homo sap
								W31782	Human parathyroidin mRNA, complete cds.
45	CATGTCCTGCCCCAT	H883029	3	19	14	27	16	M24398	
46	CATGAAGTGGCAAGA	H47683	0	16	0	0	0		Human defensin 6 (HD-6) gene, complete cds.
47	CATGGGTATTAAACCA	H708358	0	16	0	0	0	U33317	Homo sapiens defensin 6 mRNA, complete cds.
								M98331	Human mRNA for T cell receptor V beta 14 CDR3, par
48	CATGGGCTACACCTT	H684312	2	16	0	2	1	D32027	A1225F Homo sapiens cDNA clone A1225 similar to Mi
								T11701	Human fetal brain cDNA 5'-end GEN-051G02.
49	CATGAGGGTGTTC	H175870	1	15	0	0	0	D51783	Human mRNA for dipeptidase.
50	CATGCAAGGACCAGC	H272467	0	13	0	2	0	D13138	Homo sapiens (clones MDP4, MDP7) microsomal dipept
									RDP=renal dipeptidase (human, kidney, Genomic, 357
									Human alpha-1 collagen gene, 3' end with polyA sit
51	CATGTGGAATGACC	H950498	0	13	0	167	0	M10629	ym17e04.s1 Homo sapiens cDNA clone 47962 3' simila
52	CATGATCCGCCTGCC	H219514	1	13	3	4	1	H11641	yq51a09.s1 Homo sapiens cDNA clone 199288 3' simil
								R95667	
53	CATGTCCCGTACAC	H875282	1	13	0	0	1		
54	CATGATGTAAAAAAT	H241665	0	11	0	12	14	M74090	Human TB2 gene mRNA, 3' end.

										J03801	Human lysozyme mRNA, complete cds with an Alu repe
										M19045	Human lysozyme mRNA, complete cds.
55	CATGCCAGCCCCGTC	H337244	0	11	0	0	0	0			Human I-8D gene from interferon-inducible gene fam
56	CATGACCAATTCTCT	H85882	0	10	1	26	3			X57351	Human interferon-inducible mRNA (cDNA 1-8).
										X02490	Human interferon-inducible mRNA (cDNA 1-8).
57	CATGAGGACCATCGC	H165175	0	10	0	0	0	0			Human SPARC/osteonectin mRNA, complete cds.
58	CATGATGTGAAGAGT(A)	H243747	0	10	0	165	0			J03040	Human RNA fragment from patients with Crohn's dise
59	CATGCAGTTGGTTGT	H310975	0	10	6	7	4			U55217	Human RNA fragment from patients with Crohn's dise
60	CATGGCCCTCTGCCA	H613862	0	10	2	15	7				Human chaperonin-like protein (HTR3) mRNA, complet
61	CATGTTAGATAAGCA	H992010	0	10	3	3	6			M94083	Human chaperonin protein (Tep20) gene complete cds
										L27706	Human chaperonin protein (Tep20) gene complete cds

Transcripts increased in both colon primary tumors and colon cancer cell lines compared to normal colon (47 genes)

NC: Normal Colon  
 TU: Colon Primary Tumor  
 CL: Colon Cancer Cell Line  
 PT: Pancreatic Primary Tumor  
 PC: Pancreatic Cancer Cell Line

#	Tag Sequence	Tag Number	NC	CT	CL	PT	PC	Accession	Gene Name
1	CATGGCAGCCATCCG	H599350	87	180	230	72	138	U14969	Human ribosomal protein L28 mRNA, complete cds.
2	CATGATGGCTGGTAT	H239533	52	153	318	80	294	X17206	Human mRNA for LLRep3.
3	CATGCCCGTCCGAA	H355689	87	142	246	178	250	X64707	H.sapiens BBC1 mRNA
4	CATGAGGCTACGGAA	H171113	44	117	167	86	147	X56932	H.sapiens mRNA for 23 kD highly basic protein
5	CATGAGCACCTCCAG	H148949	42	116	197	103	190	Z11692	H.sapiens mRNA for elongation factor 2.
6	CATGCTGGGTTAATA	H502724	29	115	160	75	134	M81757	H.sapiens S19 ribosomal protein mRNA, complete cds
7	CATGGGATTTGGCCT	H671654	55	108	222	73	185	M17887	Human acidic ribosomal phosphoprotein P2 mRNA, com
8	CATGTACCATCAATA	H807748	46	107	98	64	189	X53778	H.sapiens hng mRNA for uracil DNA glycosylase.
								J02642	Human glyceraldehyde 3-phosphate dehydrogenase mRNA
9	CATGTGGGCAAGCC	H959498	51	103	156	45	152	Z11531	H.sapiens mRNA for elongation factor-1-gamma.
								M55409	Human pancreatic tumor-related protein mRNA, 3' en
10	CATGAATCCTGTGGA	H55227	30	95	102	48	156	Z28407	H.sapiens mRNA for ribosomal protein L8.
11	CATGGGACCACTGAA	H660601	36	92	114	43	63	X73460	H.sapiens mRNA for ribosomal protein L3.
12	CATGAGGGCTTCCAA	H174037	47	91	167	91	155	M73791	Human novel gene mRNA, complete cds.
								M64241	Human Wilms' tumor-related protein (QM) mRNA, comp
								S35960	laminin receptor homolog {3' region} [human, mRNA
13	CATGAAGGTGGAGGA	H44683	48	91	182	113	215	X80822	H.sapiens mRNA for ORF.
14	CATGTGCAGTTTTC	H935680	45	87	105	61	122	X03342	Human mRNA for ribosomal protein L32
15	CATGTCAGATCTTG	H861056	37	81	93	50	92	M58458	Human ribosomal protein S4 (RPS4X) isoform mRNA, c
								M22146	Human scar protein mRNA, complete cds.
16	CATGTGGTGTGAGG	H965603	42	79	83	55	250	X69150	H.sapiens mRNA for ribosomal protein S18.
								L06432	Homo sapiens 18S ribosomal protein (HKE3) mRNA seq
17	CATGCCTAGCTGGAT	H379369	28	77	80	46	143	Y00052	Human mRNA for T-cell cyclophilin.
18	CATGCTTGGGTTTG	518912	0	73	42	0	0	X07868	Human DNA for insulin-like growth factor II (IGF-2);
19	CATGTCCTCACCTG	H482584	12	72	41	34	50	U16811	Human Bak mRNA, complete cds.

		H507577	17	65	116	48	103	D14530	Human homolog of yeast ribosomal protein S28, comp
20	CATGCTGTTGGTGAT	H416261	28	62	183	55	94	X73974	H.sapiens HRPL4 mRNA.
21	CATGCCCGGGAACAC	H274492	9	60	73	55	119	D23661	Human mRNA for ribosomal protein L37, complete cds
22	CATGCAATAAATGTT	H79065	15	57	82	42	118	L06505	Human ribosomal protein L12 mRNA, complete cds.
23	CATGACATCATCGAT	H1000193	12	56	154	49	99	M17886	Human acidic ribosomal phosphoprotein P1 mRNA, com
24	CATGTTCAATAAAAA	H528694	24	56	71	24	146	X63527	Human mRNA for ribosomal protein L19.
25	CATGGAACACATCCA	H998030	7	55	78	35	77	M24194	H.sapiens mRNA for ribosomal protein homologous to chicken B complex
26	CATGTTATGGGATCT		18	53	50	19	61	U14967	Human ribosomal protein L21 mRNA, complete cds.
27	CATGGCATAATAGGT		23	50	103	49	120	X55954	Human mRNA for HL23 ribosomal protein homologue.
28	CATGATTCTCCAGTA	H253260						X52839	Human mRNA for ribosomal protein L17.
29	CATGACTCCAAAAAA	H119809	15	49	64	21	64	H38868	yp61a04.r1 Homo sapiens cDNA clone 191886 5' simil
								H71935	ys15f12.r1 Homo sapiens cDNA clone 214895 5'.
								Z43914	H. sapiens partial cDNA sequence; clone c-1od03.
								T48545	hbc3221 Homo sapiens cDNA clone hbc3221 5'end.
								X04347	Human liver mRNA, fragment DNA binding protein UPI
								X00910	Human mRNA for IGF-II precursor (insulin-like grow
30	CATGCTGTTGATTGC	H507455	9	44	54	22	40	X61156	Human mRNA for laminin-binding protein.
31	CATGTACAAAATCGA	802871	0	42	20	0	0	X03799	H.sapiens mRNA for laminin-binding protein mRNA
32	CATGGAAAAATGGTT	H524524	14	41	81	15	57	U02032	Human colon carcinoma laminin-binding protein mRNA
								U14970	Human ribosomal protein L23a mRNA, partial cds.
								X58965	Human ribosomal protein S5 mRNA, complete cds.
33	CATGAAGAAGATAGA	H33331	9	39	69	30	56	X58965	Human ribosomal protein S5 mRNA, complete cds.
34	CATGCCCTCGAGATC	H125661	5	29	25	25	38	M36981	H.sapiens RNA for nm23-H2 gene.
35	CATGACTGGGTCTAT							L16785	Human putative NDP kinase (nm23-H2S) mRNA, complet
								L10376	Homo sapiens c-myc transcription factor (puf) mRNA
								S80520	Human (clone CTG-B33) mRNA sequence.
36	CATGCAGCTCACTGA	H302367	9	29	40	27	31	M77349	CAG-isl 7 (trinucleotide repeat-containing sequenc
								X58536	Human transforming growth factor-beta induced gene
37	CA1GGTGTGTTTGTGA	H1769020	0	24	15	22	8	X00497	Human mRNA for HLA class I locus C heavy chain.
38	CA1GGTGGCTGAGC	H760291	0	22	17	44	18	X16934	Human mRNA for HLA-DR antigens associated invarian
39	CA1GGTTCACATTAG	H174461	3	22	25	141	10	Y00345	Human hB23 gene for B23 nucleophosmin.
40	CATGTGAAATAAAAC	H918273	2	18	37	8	22	X81005	Human mRNA for polyA binding protein.
41	CATGAAAAGAAACTT	H2056	1	16	27	11	25	D28137	H.sapiens HCG IV mRNA.
42	CATGTGCTGCCTGTT	H948604	1	15	16	11	3		Human mRNA for BST-2, complete cds.
									Soares senescent fibroblasts NBHSF Homo sapiens cDNA clone
									324128 3'.
									H.sapiens DNA for orphan TCR V-beta segment (allele
43	CATGCTGATGGCAGA	H495251	0	14	15	8	6	W46476	
								X72718	

44	CATGACTCGCTCTGT	H121311	0	12	16	5	7	H121311	Soares fetal heart NbHH19W Homo sapiens cDNA clone 342926
								AA305589	3'
45	CATGGCCCCAAGGACC	H610466	0	12	19	82	17	X33416	EST176663 Colon carcinoma (Caco-2) cell line II Homo sapiens cDNA 5' end
46	CATGATCTTGTTACT	H229106	0	11	28	67	0	X02761	Human mRNA for actin-binding protein (filamin) (AB
47	CATGAAGCTGCTGGA	H40571	0	10	17	6	6	Z26305	Human mRNA for fibronectin (FN precursor).
									H.sapiens isoform 1 gene for L-type calcium channe

# Transcripts increased in only colon cancer cell lines compared to normal colon (181 genes)

NC: Normal Colon

TU: Colon Primary Tumor

CL: Colon Cancer Cell Line

PT: Pancreatic Primary Tumor

PC: Pancreatic Cancer Cell Line

#	Tag Sequence	Tag Number	NC	TU	CL	PT	PC	Accession	Gene Name
1	CATGTGTGTTGAGAG	H978825	71	79	487	136	412	X16869	Human mRNA for elongation factor 1-alpha
2	CATGCCGAGGAAGG	H615043	72	66	265	105	125	X53505	Human ribosomal protein S12.
3	CATGCCAACCATCCA	H263478	137	83	245	36	502	X12883	Human cytokeratin 18.
4	CATGCACAAACGGTA	H278636	63	53	201	74	179	L19739	Homo sapiens metalloproteinase (MPSI)
5	CATGAAAAAATAAAA	H1	31	48	186	66	102	X83412	H.sapiens BI mRNA for mucin.
								Z32564	H.sapiens FRGAMMA mRNA (819bp) for folate receptor
								X76180	H.sapiens mRNA for lung amiloride sensitive Na+ ch
								U08470	Human FR-gamma' mRNA, complete cds.
								U08471	Human folate receptor 3 mRNA, complete cds.
								S64030	Human L41 ribosomal protein
6	CATGTTGGTCTCTCTG	H1027448	115	128	179	104	358	T91925	ye0202.r1 Homo sapiens cDNA clone 116571 5'.
7	CATGTCCTCCATACCC	H906438	0	0	176	48	0	X66699	H.sapiens ribosomal protein L37a.
8	CATGAAGACAGTGGC	I133979	59	61	172	55	252	M60854	Human ribosomal protein S16
9	CATGCCGTCCTCAAGG	I1374027	50	39	138	60	108	M92381	Human thymosin beta 10
10	CATGGGGGAAATCGC	H696375	90	90	136	203	231	X69181	H.sapiens mRNA for ribosomal protein L31.
11	CATGAAGGAGATGGG	H41531	30	37	133	38	161	U14968	Human ribosomal protein L27a
12	CATGGAGGGAGTTTC	H567488	38	53	112	65	142	X79234	Human ribosomal protein L11.
13	CATGCCGCTGGTTCCA	H424694	42	64	111	53	49	J03537	Human ribosomal protein S6
14	CATGGCCGTGTCCGC	H618199	56	39	109	28	120	U58682	Human ribosomal protein S28 mRNA, complete cds
15	CATGGACGACACGAG	H549145	32	59	105	44	70	X52839	Human mRNA for ribosomal protein L17
16	CATGTACCCACACACC	H857362	36	48	103	44	65	U12465	Human ribosomal protein L35
17	CATGCCGCCCGGCT	H416106	39	43	90	52	184	M17885	Human acidic ribosomal phosphoprotein P0
18	CATGCTCAACATCTC	H473448	27	41	89	27	145	M23725	Human M2-type pyruvate kinase mRNA, complete cds.
19	CATGTGGCCCCACCC	H955718	20	30	80	46	55	M26252	Human TCB gene encoding cytosolic thyroid hormone-
20	CATGCCCTGGGTCT	H359102	34	49	78	92	145	M11147	Human ferritin L chain

21	ICATGAGCATCTCCAG	H150997	0	0	77	0	0	H09058	yl96f1.l.r1 Homo sapiens cDNA clone 45943 5'.
								Z44640	H. sapiens partial cDNA sequence; clone c-26b05.
								N75111	yz29e01.r1 Homo sapiens cDNA clone 284472 5'.
22	CATGGCCTGTATGAG	H621369	24	32	77	33	99	M31520	Human ribosomal protein S24 mRNA.
23	CATGAGCTCTCCCTG	H161624	33	39	76	21	67	X53777	Human L23 mRNA for putative ribosomal protein.
									gb AA223340 AA223340 Homo sapiens cDNA clone 650651 3' similar to
24	CATGCCAGGAGGAAT	H338081	27	12	74	23	87	AA223340	gb:Y00371 rna1 HEAT SHOCK COGNATE 71 KD PROTEIN (HUMAN)
25	CATGGGCAAGCCCCA	H672342	30	55	72	27	61	U12404	Human Csa-19
26	CATGAGGAAAGCTGC	H163999	31	42	70	32	146	F16378	H. sapiens EST sequence (135-18) from skeletal muscle
27	CATGAACGGGCCAA	H26261	29	46	69	54	79	Z23063	Homo sapiens macrophage migration inhibitory factor
28	CATGCCAGAACAGAC	H335945	23	39	66	42	148	X79238	H. sapiens ribosomal protein L30.
29	CATGGCCGCCATCTC	H615736	7	10	65	10	22	U55017	Human transketolase (TKT)
30	CATGGTGTAAACAG	H769045	16	19	65	17	76	L25899	Human ribosomal protein L10
31	CATGCCTCGGAAAT	H383489	9	13	64	23	46	Z26876	H. sapiens ribosomal protein L38.
32	CATGAGGTCCTAGCC	H177610	15	27	63	43	41	X06547	Human class Pi glutathione S-transferase
33	CATGGTTCCTGGCC	H775658	31	26	63	32	96	X65923	H. sapiens fau mRNA.
34	CATGTAAGGAGCTGA	H796831	32	58	62	42	68	X77770	H. sapiens RPS26
35	CATGAACATAAAAAA	H28673	7	14	60	17	39	W52460	zc45e1.l.r1 Soares senescent fibroblasts NbHSF Homo
								N92893	zb71h03.s1 Homo sapiens cDNA clone 309077 3'.
36	CATGATTTGTCCCA	H260949	17	13	57	9	91	X14957	Human hmg1 mRNA for high mobility group protein I.
37	CATGATAATCTTTG	H200576	13	27	53	30	69	U14973	Human ribosomal protein S29
38	CATGCCCCAGCCAGT	H348756	18	23	53	5	85	U14990	Human XP1PO ribosomal protein S3 (rps3)
39	CATGGGAGTGGACAT	H667269	15	13	49	13	45	L11566	Homo sapiens ribosomal protein L18 (RPL18)
40	CATGTAAAAAATAAAA	H786433	13	8	48	10	26	H08238	yl87a01.r1 Homo sapiens cDNA clone 44932 5'.
41	CATGGTGTTCACAA	H769605	19	21	48	21	47	X79239	H. sapiens ribosomal protein S13.
42	CATGGCCAGCCCA	H608595	6	21	47	11	15	U31657	Human unknown protein mRNA, partial cds.
								H41030	yn92a10.r1 Homo sapiens cDNA clone 175866 5'.
43	CATGGGCTCCCACTG	H685384	14	24	47	23	15	M16660	Human 90-kDa heat-shock protein
44	CATGTCAACTTCTGG	H853983	0	0	46	2	0	N57419	yw82e04.r1 Homo sapiens cDNA clone 258750 5' simil
45	CATGGATGCTGCCAA	H583573	6	12	46	27	18	X59357	Human mRNA for Epstein-Barr virus small RNAs (EBER)
								L21756	Homo sapiens acute myeloid leukemia associated protein
								D17652	Human mRNA for HBp15/L22, complete cds.
46	CATGAATAGGTCCAA	H51925	13	31	46	47	53	M64716	Human ribosomal protein S25
47	CATGGCTTTAAGGA	H655115	8	26	45	22	63	L06498	Homo sapiens ribosomal protein S20 (RPS20)
48	CATGAATGCAGGCAG	H585533	2	12	44	6	27	M61831	Human S-adenosylhomocysteine hydrolase (AHCY)



		H610939	8	18	43	0	22	Z21507	Human elongation factor I delta (EF 1delta)
49	CATGGCCAGCTGGA	H610939	6	6	42	8	18	M13932	Human ribosomal protein S17 mRNA
50	CATGGGCGCGTTTCG	H678334	14	26	42	15	42	M10036	Human triosephosphate isomerase
51	CATGTGAGGGAATAA	H928269	14	24	42	35	49	K00558	human alpha-tubulin
52	CATGTGTACCTGTAA	H968173	8	7	41	12	87	L19527	Homo sapiens ribosomal protein L27 (RPL27)
53	CATGGGCAAGAAGAA	H672265	6	14	40	14	15	X63237	H.sapiens Uba80 mRNA for ubiquitin.
54	CATGAACCTAACAAAA	H28737	0	0	38	0	9		Unknown
55	CATGTATACGCTCAG	H837237	7	17	38	14	42	X69391	H.sapiens ribosomal protein L6.
56	CATGTACAAGAGGAA	H803369	8	17	38	12	25	H11182	ym14a02.r1 Homo sapiens cDNA clone 47866 5'
57	CATGGTTAACGTCCTC	H770486	8	17	38	12	25	T40302	ya31g04.r5 Homo sapiens cDNA clone 62262 5'
								T89480	yd98a05.r1 Homo sapiens cDNA clone 116240 5'
								H01362	yi99c06.r1 Homo sapiens cDNA clone 147370 5'
58	CATGGAGACTCCTGC	H558943	13	12	38	32	10	H94371	yw34e05.r1 Homo sapiens cDNA clone 256064 5'
59	CATGATCCACATCGC	H217399	3	10	37	10	14	T49412	ya75b09.r1 Homo sapiens cDNA clone 67481 5'
								T51058	yb55a12.r1 Homo sapiens cDNA clone 75070 5'
								X07270	Human heat shock protein hsp86.
60	CATGGAAGCTTTGCA	H534522	11	13	37	14	25	M91670	Human ubiquitin carrier protein (E2-EPF)
61	CATGCTGGCGAGCGC	H501287	2	9	36	3	18	X74070	H.sapiens transcription factor BTF 3.
62	CATGCTGAGACAAAG	H493633	13	8	36	8	26	V00599	Human beta-tubulin
63	CATGAACGAGCTCGT	H24951	7	13	35	22	40	X84694	H.sapiens mRNA for elongations factor Tu-mitochondria
64	CATGGCATAGGCTGC	H602783	9	16	35	2	17	L38995	Homo sapiens nuclear-encoded mitochondrial elongation factor
								S75463	Homo sapiens nuclear-encoded mitochondrial elongation factor (human P43=mitochondrial elongation factor homolog)
								H48893	yq80b12.r1 Homo sapiens cDNA clone 202079 5'
65	CATGCATCTTCACCA	H319302	12	14	35	9	16	X71973	H.sapiens GPx-4 mRNA for phospholipid hydroperoxidase
66	CATGGCCTGCTGGGC	H621035	10	5	32	18	107	M95787	Human 22kDa smooth muscle protein (SM22)
67	CATGACAGGCTACGG	H76231	0	5	31	64	0	H80294	yu59g01.s1 Homo sapiens cDNA clone 230448 3'
68	CATGGAAATGTAGA	H528067	5	12	31	14	25	R74294	yi57f06.r1 Homo sapiens cDNA clone 143363 5'
								L36035	Human 4E-binding protein 1
69	CATGGAAGCCAGCCA	H533798	1	3	30	9	11	F17005	H.sapiens EST sequence (011-T1-18) from skeletal muscle
70	CATGTTACCATATCA	H988366	10	28	30	19	86	H10519	y190g04.r1 Homo sapiens cDNA clone 45563 5'
71	CATGTTGCTCACAAA	H1023249	1	2	29	1	2		Unknown
72	CATGTCCCGCTCGA	H874103	0	6	29	0	0	X04409	Human coupling protein G(s) alpha-subunit
73	CATGATTAAACAAGC	H246019	8	9	29	25	26	X56998	Human Uba52 adrenal mRNA for ubiquitin-52 amino acid
74	CATGCAGATCTTTGT	H298495	2	7	28	8	24	F19234	H.sapiens EST sequence (005-X3-16) from skeletal m
75	CATGGTTCGTGCCAA	H777109	9	28	28	17	46	X52317	Human histone H2A.Z.
76	CATGGACGTGTGGC	H552683	3	4	27	2	16		

77	CATGCTAAAAA	H458753	4	8	27	19	8	M33680	Human 26-kDa cell surface protein TAPA-1
78	CATGGGGTTTATT	H704500	4	1	27	6	18	L28809	Homo sapiens dbpB-like protein
79	CATGCGGATCACCGG	H363799	7	9	27	7	15	M29536	Human translational initiation factor 2 beta subunit
80	CATGGCACAAGAGA	H594051	6	9	26	7	29	W07137	za92a1.l.r1 Soares fetal lung NbHL19W Homo sapiens
								D20503	Human HL60 3'directed Mbol cDNA, HUMGS01477, clone
								N91592	Soares fetal lung NbHL19W Homo sapiens cDNA clone 303055 3'
								H83884	yv84c07.s1 Homo sapiens cDNA clone 249420 3' similar to contains Alu repetitive element;
81	CATGCTCTACCCAC	H908373	7	11	26	11	13	Z22572	H.sapiens CDE1 binding protein mRNA.
								L09209	Homo sapiens amyloid protein homologue mRNA, compl
								L19597	Human binding protein mRNA, partial cds.
								S60099	APPH=amyloid precursor protein homolog (human, pla
82	CATGGTTTCCCAAG	H783697	1	0	25	3	0	W07587	zb06f02.r1 Soares fetal lung NbHL19W Homo sapiens
								N28502	yx36f06.r1 Homo sapiens cDNA clone 263843 5'
								N35630	yx62a03.r1 Homo sapiens cDNA clone 266284 5'
83	CATGCCCTGTCCAGCC	H388426	2	3	25	3	13	Z40265	H. sapiens partial cDNA sequence; clone c-1xe03.
								W02723	zc65e03.s1 Soares fetal heart NbHL19W Homo sapiens
								N24893	yx99h09.s1 Homo sapiens cDNA clone 269921 3'
								N32178	yy25b09.s1 Homo sapiens cDNA clone 272249 3'
84	CATGTCATCATCTGA	H865503	5	15	25	5	7	H21873	y34b10.s1 Homo sapiens cDNA clone 160123 3' simil
								H26394	y48e12.s1 Homo sapiens cDNA clone 161518 3' simil
								H69857	yr88d02.s1 Homo sapiens cDNA clone 212355 3' simil
								H70714	yu69b11.s1 Homo sapiens cDNA clone 239037 3' simil
85	CATGCCCTGCCTTGT	H358783	5	8	25	16	31	X55110	Human mRNA for neurite outgrowth-promoting protein
86	CATGGCCGGGCCCTC	H617048	1	1	24	0	1	X03168	Human mRNA for S-protein.
87	CATGTTGCTCAAAA	H1023233	2	1	24	2	2	AA143561	zo32d09.s1 Stralagene colon (#937204) Homo sapiens cDNA clone 588593
								AA152342	3' similar to contains LTR7.t1 LTR7 repetitive element
								AA115727	zo01g11.s1 Stralagene colon (#937204) Homo sapiens cDNA clone 566468
								AA115727	3' similar to contains LTR7.t1 LTR7 repetitive element ;
								AA115727	z186h11.s1 Stralagene colon (#937204) Homo sapiens cDNA clone 511557
88	CATGCAAAATCAGGA	H262987	6	2	24	5	15	R76502	3' similar to contains LTR7.t1 LTR7 repetitive element
								T32681	y161f09.r1 Homo sapiens cDNA clone 143753 5'.
								T34662	EST52915 Homo sapiens cDNA 5' end similar to None.
								H04634	EST72468 Homo sapiens cDNA 5' end similar to None.
89	CATGGAAGATGTGGG	H533435	1	5	23	4	7	H04634	y149h03.r1 Homo sapiens cDNA clone 152117 5'.

90	CATGGTGCTCATTTCA	H761150	0	8	23	6	4	F00364	H. sapiens partial cDNA sequence; clone 76D12; ver
								H01503	y121c05.s1 Homo sapiens cDNA clone 149384 3'.
								H84813	yv86c02.s1 Homo sapiens cDNA clone 249602 3' simil
								H84956	yv88f07.s1 Homo sapiens cDNA clone 249829 3' simil
								L38961	Homo sapiens putative transmembrane protein (B5)
								J04026	Human thioredoxin (TXN) mRNA
91	CATGGCTTTACTTTG	H654464	4	5	23	9	5	D11078	Human RGH2 gene.
92	CATGTTTTCTGAAAA	H1046401	6	13	23	10	10	X53279	Human mRNA for placental-like alkaline phosphatase
93	CATGTTGCTCACACA	H1023250	1	4	22	0	4	M77836	Human pyrroline 5-carboxylate reductase mRNA.
94	CATGGAATTCACG	H589267	0	0	22	0	19	X07674	Human glutamate dehydrogenase
95	CATGAGGAGGAGGC	H166539	2	3	22	2	4	Y00433	Human mRNA for glutathione peroxidase
96	CATGGCTTAACCTGG	H651359	3	4	22	2	19	X67951	H. sapiens mRNA for proliferation-associated gene
97	CATGCTCTTCGAGAA	H490889	4	8	22	27	6	U38846	Human stimulator of TAR RNA binding (SRB)
98	CATGAGAACAAAACC	H132098	1	7	21	9	24	D16933	Human HepG2 3' region cDNA, clone hmd4f11.
99	CATGCCACGGAGAA	H346761	3	3	21	2	107	U42376	Human retinoic acid induced RIG-E
									Unknown
100	CATGCACCTTCAAGG	H294155	0	3	20	47	1	F17524	H. sapiens EST sequence (012-T2-32) from skeletal m
101	CATGGCGGAGAGAG	H631331	2	3	20	4	22	W52942	Unknown
102	CATGTTACCTCCTTC	H989024	4	7	20	3	7	R21316	zc031h05.r1 Soares parathyroid tumor NbH19A Homo sap
103	CATGACTCTGCCAAG	H122449	4	7	19	12	7	X00566	yv48f11.r1 Homo sapiens cDNA clone 35917 5' simila
104	CATGTCAGATGGCGT	H861095	1	6	19	5	3	M80244	Human lipoprotein apoA1.
105	CATGGGCTTTTCT	H167936	1	3	19	0	0	H27927	Human E16 mRNA
106	CATGTGGACGCGCTG	H951912	0	0	19	6	5	X57959	y158c11.s1 Homo sapiens cDNA clone 162452 3' simil
107	CATGCTGCTGCCCTG	H386904	0	5	18	18	15	AA299898	H. sapiens ribosomal protein L7.
108	CATGGCCACACCCCA(C)	H607318	2	6	18	5	20	U09510	EST12509 Uterus tumor 1 Homo sapiens cDNA 5' end
109	CATGATTATTTTCT	H249854	2	3	18	5	15	X76013	Human glycyl-tRNA synthetase
110	CATGGAACCTGGGA	H529899	2	7	18	8	17	W16529	H. sapiens QRSHs mRNA for glutamyl-tRNA synthetase
111	CATGGGCTGATGTGG	H686319	3	5	18	4	4	W35192	zb10a11.r1 Soares fetal lung NbHL19W Homo sapiens
112	CATGTCAATAAGAA	H855049	3	10	18	0	5	W52451	zc70b05.r1 Soares fetal heart NbHH19W Homo sapiens
113	CATGAAAGTGAAGAT	H11785	0	7	17	0		D38251	zc45d09.r1 Soares senescent fibroblasts NbHSF Homo
								D52570	Human mRNA for RPB5 (XAP4)
								D52758	Human fetal brain cDNA 5'-end GEN-081G12.
114	CATGCACGCGCTCAA	H288373	0	1	17	0	31	D52758	Human fetal brain cDNA 5'-end GEN-087A08.
115	CATGAACTAATACTA	H28872	1	6	17	13		D55953	Human fetal brain cDNA 5'-end GEN-407H12.
								M22490	Human bone morphogenetic protein-2B (BMP-2B)
116	CATGCTGTACCTGGA	H504187	1	0	17	12	6		

117	CATGCGACCCACGC	H398663	2	6	17	48	0	M12529	Human apolipoprotein E
118	CATGTAGAAAATAA	H819213	0	1	16	2	7	X16539	H.sapiens RNA for neuroleukin gene.
119	CATGATCTTGAAAGG	H228867	0	0	16	5	3	M27691	Human transactivator protein (CREB) mRNA, complete
120	CATGCAGCTGGCCAT	H302741	0	1	16	14	0	X53743	H.sapiens NAP (nucleosome assembly protein)
121	CATGATCTTGAAAGG	H228867	0	0	16	5	3	Z26328	H.sapiens mRNA for fibulin-1 C.
121	CATGATCTTGAAAGG	H228867	0	0	16	5	3	Z26328	H.sapiens partial cDNA sequence; clone HEC059
122	CATGGTGGAGGTGCG	H762554	2	10	16	3	5	U22055	H.sapiens partial cDNA sequence; clone HEC059
123	CATGGTGGACCCAA	H762197	1	3	15	7	10	R91724	Human 100 kDa coactivator mRNA
								W51770	yp98e02.r1 Homo sapiens cDNA clone 195482 5' simil
								N42086	zc48a02.r1 Soares senescent fibroblasts NbHSF Homo
124	CATGGAGCAGCTGGA	H561787	0	5	15	2	4	R80990	yy05b03.r1 Homo sapiens cDNA clone 270317 5'
								R95036	yi94c02.r1 Homo sapiens cDNA clone 146882 5'
125	CATGGCGGAGGGCT	H633002	1	6	15	8	7	F16507	yq44f01.r1 Homo sapiens cDNA clone 198649 5' simil
								T50201	H.sapiens EST sequence (147-09) from skeletal musc
								S85655	yb77h03.r1 Homo sapiens cDNA clone 77241 5' simila
126	CATGATTGGCTTAAA	H256497	1	8	15	0	16	M38188	Human prohibitin
127	CATGGAAAAATTAA	H524541	0	3	15	4	0	Y00711	Human unknown protein from clone pHGR74 mRNA, comp
128	CATGGATCACAGTTT	H577840	0	5	15	0	0	D83174	Human lactate dehydrogenase B (LDH-B).
129	CATGAGCCTTTGTTG	H155632	1	2	15	23	5	X70940	Human collagen binding protein 2.
130	CATGCTGCACCTCC	H910430	0	0	15	0	2	T30623	H.sapiens elongation factor 1 alpha-2.
131	CATGAACAGAAGCAA	H18469	0	2	15	3	11	C01011	EST19638 Homo sapiens cDNA 5' end similar to None.
								AA111865	HUMGS0004747, Human Gene Signature, 3'-directed cDNA
								W56516	sequence.
								W30299	zm62d06.s1 Stralagene fibroblast (#937212) Homo sapiens cDNA clone
132	CATGTGTTTCAGGACC	H980130	1	1	14	5	11	H50265	S30219 3'
								W01702	zd16c08.r1 Soares fetal heart NbHH19W Homo sapiens
133	CATGTAGATAATGCC	H822331	1	4	14	6	14	W04495	yo77d04.r1 Homo sapiens cDNA clone 183943 5' simil
								W23328	yo28e02.r1 Homo sapiens cDNA clone 179234 5'.
								D11838	za37a06.r1 Soares fetal liver spleen 1NFLS Homo sa
								X75598	za58b10.r1 Soares fetal liver spleen 1NFLS Homo sa
134	CATGCTTAATCCTGA	H508767	0	6	14	6	12	H1838	zc71g11.s1 Soares fetal heart NbHH19W Homo sapiens
135	CATGGGCAGAGGACC	H673954	0	6	14	5	11	T35470	Human HepG2 3'-directed Mbol cDNA, clone hm02e09.
136	CATGTGACTGAAGCC	H925194	0	5	14	3	0	T35536	H.sapiens nm23H1 gene.
									EST85850 Homo sapiens cDNA 5' end similar to None.
									EST86951 Homo sapiens cDNA 5' end similar to None.

									T35545	EST87066 Homo sapiens cDNA 5' end similar to None.
137	CATGGATAGTTGTGG	H576495	0	1	14	2	1		H01694	yj3gl1.s1 Homo sapiens cDNA clone 150596 3'.
									N78851	zb17d08.s1 Homo sapiens cDNA clone 302319 3'.
									N78931	za2ht06.s1 Homo sapiens cDNA clone 300059 3'.
138	CATGGTGGTGGACAC	H765573	1	4	13	6	13		H90469	yo1e06.r1 Homo sapiens cDNA clone 241474 5' simil
									R76765	yf63g01.r1 Homo sapiens cDNA clone 143952 5' simil
									T35045	EST79335 Homo sapiens cDNA similar to None..
139	CATGTGGGGTACCCTT	H961304	0	6	13	2	9		yo31a05.r1 Homo sapiens cDNA clone 179504 5'.	
									w46469	zc32c05.r1 Soares senescent fibroblasts NbHSF Homo
									w51800	zc48e04.r1 Soares senescent fibroblasts NbHSF Homo
									R33196	yh77f08.r1 Homo sapiens cDNA clone 135783 5'.
									J04799	Human prothymosin-alpha
140	CATGTTCAATTATAAT	H1003313	1	10	13	8	10		D80012	Human KIAA0190 protein
141	CATGCTTCTGTGTACT(T)	H515821	0	5	13	8	12		U02389	Human hLON ATP-dependent protease mRNA
142	CATGACTGCCGAAGT	H125315	1	5	13	2	5		T29819	EST96617 Homo sapiens cDNA 5' end similar to ATP-d
									X14850	Human histone H2A.X.
143	CATGGAAGAAGCTGA	H526495	1	3	13	1	6		J04088	Human DNA topoisomerase II (top2) mRNA
144	CATGCAACTCTATGG	H269775	0	1	13	1	2		K01891	Human beta globin retrovirus-like repetitive element
145	CATGAAATTTGGTGC	H16303	0	0	13	0	0		I18396	EST28e05 Homo sapiens cDNA clone 28e05
									X74796	H.sapiens p85Mcm mRNA.
146	CA TGC GCAC TTACT	H496114	1	2	13	1	8		D28480	Human mRNA for hMCM2, complete cds.
									D55716	Human B lymphoma mRNA for Plcdc47, complete cds.
147	CATGAATATTGAGAA	H53129	0	5	13	6	11		T30327	EST14849 Homo sapiens cDNA 5' end similar to None.
									T34394	EST66942 Homo sapiens cDNA 5' end similar to None.
									T47475	yb14c03.r1 Homo sapiens cDNA clone 71140 5'.
									T50289	yb14h08.r1 Homo sapiens cDNA clone 71199 5'.
									Unknown	Unknown
148	CATGTCGGCGGCGC	H890535	0	1	13	2	1		H59914	Unknown
149	CATGGGGGCAGCCG	H697495	0	2	13	2	7		U33818	Human inducible poly(A)-binding protein
150	CATGCCAAGAAAGAA	H329737	0	6	12	4	4		D16891	Human HepG2 3' region cDNA, clone hind2cl1.
151	CATGTTTTTGATAAA	H1048113	0	5	12	4	12		M29882	Human apolipoprotein A-II
152	CATGTGTGGAGAGCC	H977034	0	0	12	0	0		Z49216	H.sapiens mitoxantrone-resistance associated mRNA.
153	CATGCCACCGTTAG	H145789	0	5	12	5	4		Unknown	Unknown
154	CATGAATTCCTAA	H63325	0	1	12	1	1		Unknown	Unknown
155	CATGGACCTCCGGC	H548203	0	0	12	0	0		M93651	Human set gene
156	CATGTGAATCTGGGT	H921067	0	2	11	7	8			

157	CATGTCCTCTCCAC	H884181	0	5	11	14	8	X15804	Human alpha-actinin.
158	CATGTACTGTCTAC	H843485	0	4	11	2	3	T19569	609F Homo sapiens cDNA clone 609 similar to SET protein
159	CATGACGTTCTCTC	H114144	0	0	11	1	17	Z36249	HHEA18W H. sapiens partial cDNA sequence; clone HEA18W;
160	CATGCCCTGAGTCAG	H358581	0	0	11	0	0	AA207189	zq73e07.r1 Stratagene neuroepithelium (#937231) Homo sapiens cDNA clone 647268 5' similar to TR:E16910 E16910 ENDONUCLEASE.;
161	CATGGAATTCCTCGA	H340023	0	3	11	3	1	JN80776	z98104.s1 Homo sapiens cDNA clone 300631 3'.
								AA025809	z90d01.s1 Soares fetal heart NbHH19W Homo sapiens cDNA clone 366241 3'.
								AA279492	z585h05.s1 Soares Nb11TGBC Homo sapiens cDNA clone 704313 3'.
162	CATGGACGCCGAAC	H350274	0	1	11	6	0		Unknown
163	CATGGCGGACTGGG	H631275	0	0	11	1	0	AA098867	z484f04.s1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone 489535 3' similar to SW:A5_XENLA P28824 A5 PROTEIN PRECURSOR
164	CATGGGAACACACAG	H656453	0	1	11	0	2	R48460	yj67c12.r1 Homo sapiens cDNA clone 153814 5'.
								AA173819	zp01c02.r1 Stratagene ovarian cancer (#937219) Homo sapiens cDNA clone 595106 5'.
165	CATGTTGGGAGCCC	H1022502	0	2	11	2	1	L19183	HUMMAC30X Human MAC30 mRNA, 3' end.
								H61710	y24a07.s1 Homo sapiens cDNA clone 206196 3'.
								H77330	yu11f12.s1 Homo sapiens cDNA clone 233519 3'.
								N69482	za18d05.s1 Homo sapiens cDNA clone 292905 3'.
166	CATGGCAGACATTGA	H598335	0	7	10	4	9	H41078	yp52c11.s1 Homo sapiens cDNA clone 191060 3' simil
167	CATGCACCTTGAAAA	H294401	0	1	10	5	0	H04630	yj49g03.r1 Homo sapiens cDNA clone 152116 5'.
168	CATGGGTTGGCAGG	H719435	0	0	10	24	0	R77027	yj66e12.r1 Homo sapiens cDNA clone 144238 5'.
169	CATGTTCTCTCGGC	H1007018	0	1	10	4	12	R32331	yh68g02.s1 Homo sapiens cDNA clone 134930 3' simil
170	CATGCTGCCGAGCT	-497192	0	8	10	1	10	T86566	yd77g07.r1 Homo sapiens cDNA clone 114300 5' simil
171	CATGGTGAAAAAAA	H753665	0	2	10	3	7	S77357	transcript ch111 (human, RFI, RF48 stomach cancer c
172	CATGCTGTGCAGCA	H506149	0	6	10	6	1	M34338	Human spermidine synthase
173	CATGTAOTTTGTGG	-835515	0	1	10	0	2	U03911	Human mutator gene (hMSH2)
174	CATGATGTAGTAGTG	H242380	0	5	10	9	7	D55671	Human heterogeneous nuclear ribonucleoprotein
175	CATGGACCCACTACC	H545906	0	1	10	3	1	J03569	Human lymphocyte activation antigen 4F2 large subunit
176	CATGAAATAGGTTT	H12992	0	1	10	6	3	D53402	Human fetal brain cDNA 5'-end GEN-108D03.
								T61971	yb96f02.r1 Homo sapiens cDNA clone 79035 5'.
								D61243	Human fetal brain cDNA 5'-end GEN-171G06.
								N77240	yv44d02.r1 Homo sapiens cDNA clone 245571 5'.
177	CATGCCGGCGGTGGT	H371131	0	0	10	1	2	T35761	EST90898 Homo sapiens cDNA 5' end similar to EST c

EST40719 Homo sapiens cDNA 5' end similar to None.												
		H555168	0	8	10	3	3	T31901	EST40719 Homo sapiens cDNA 5' end similar to None.			
178	CATGGACTGAGCTTG								HSMPP41 H.sapiens mRNA for M-phase phosphoprotein, mpp4, 1523bp			
179	CATGAAACGCCCAAT	H6481	0	2	10	1	3	X98264	Unknown			
180	CATGATGAGGCCGGG	H232027	0	4	10	7	1		Human mRNA for KIAA0246 gene, partial cds			
181	CATGGCCCCACATCCG(A)	H610614	0	9	10	6	2	D87433				

Table 3 - Transcripts decreased in colon cancer

### Transcripts decreased in only colon primary tumors compared to normal colon (51 genes)

NC: Normal Colon

TU: Colon Primary Tumor

CL: Colon Cancer Cell Line

PT: Pancreatic Primary Tumor

PC: Pancreatic Cancer Cell Line

#	Tag sequence	Tag Number	NC	CT	CL	PT	PC	Accession	Gene Name
1	CATGGCTTTATTGT	H654591	184	110	185	203	111	X00351	Human mRNA for beta-actin.
2	CATGCTAGCCTCAGC	H468434	170	61	130	80	75	X04098	Human mRNA for cytoskeletal gamma-actin.
3	CATGCAAAACCATCCA	H263478	137	83	245	36	502	X12883	Human mRNA for cytokeratin 18.
4	CATGCTTCCAGCTAA	H513181	64	23	36	53	104	D00017	Human lipocortin II mRNA.
5	CATGCCCCAGTTGCT	H348922	61	27	38	37	46	X04106	Human mRNA for calcium dependent protease (small subunit)
6	CATGGATGACCCCC	H581974	53	4	42	6	32	Z65513	H.sapiens CpG island DNA genomic MseI fragment, cl
7	CATGCTGTACAGACA	H504098	50	22	26	6	32	W61077	z30402.r1 Soares fetal heart NbHH19W Homo sapiens
8	CATGCGGACTCACTG	H427848	47	15	26	18	4	D60944	Human fetal brain cDNA 5'-end GEN-141D02.
9	CATGCCCCCGCGGAA	H349801	47	10	21	15	8		Unknown
10	CATGCCCTGGAAGAGG	H387107	46	19	39	47	14	J02783	Human thyroid hormone binding protein (p55) mRNA.
11	CATGGCCTGGCCATC	H621140	46	19	24	16	20	N33042	yy05405.s1 Homo sapiens cDNA clone 270345 3'
12	CATGAGCAGGAGCAG	H150053	43	12	26	24	20	W07627	zb06a05.r1 Soares fetal lung NbHL19W Homo sapiens
13	CATGAACGTGCAGGG	H28235	42	6	57	2	10	X01630	Human mRNA for argininosuccinate synthetase.
14	CATGGCCGCCCTGCA	H615802	40	12	16	17	8	D43682	Human mRNA for very-long-chain acyl-CoA dehydrogen
15	CATGTGGGAGAGGGA	H960651	40	5	36	10	5	D29146	Human keratinocyte cDNA, clone 173.
16	CATGGCTGCCCTTGA	H648575	38	10	20	6	39	K00557	human alpha-tubulin mRNA, 3' end.
17	CATGTGCCCATCTGC	H955615	37	5	15	19	18	AA341633	AA341633 EST47188 Fetal kidney II Homo sapiens cDNA 5' end
18	CATGCGTTCCTGCGG	H456167	35	4	36	8	0	X77956	H.sapiens Id1 mRNA.
19	CATGTGCATCTGTTG	H937452	33	9	14	13	10	X87949	H.sapiens mRNA for BiP protein.
20	CATGTTGACCTCCTT	H755160	33	7	12	6	31	J04823	Human cytochrome c oxidase subunit VIII (COX8) mRNA
21	CATGTAGCTCTATGG	H826831	33	5	18	9	13	U16798	Human Na,K-ATPase alpha-1 subunit mRNA, complete c
22	CATGGTGGCTAGGG	H760267	29	7	26	19	27	R50350	gbJR50350JR50350 yJ59c04.s1 Homo sapiens cDNA clone 153030 3'
								R50013	yJ59c04.r1 Homo sapiens cDNA clone 153030 5'.
								C02981	Human Heart cDNA, clone 3NHC0642.



		EST30445 Homo sapiens cDNA 5' end similar to ubiquinol cytochrome-c reductase, 6.4 kDa.											
		Unknown											
		yr34d11.r1 Homo sapiens cDNA clone 207189 5' simil											
		zd27c08.r1 Soares fetal heart NBHH19W Homo sapiens											
		Human GTPase (rhoC) mRNA, complete cds.											
		Human mRNA for calmodulin, complete cds.											
		yy66b11.s1 Homo sapiens cDNA clone 278493 3'.											
		yi14b06.s1 Homo sapiens cDNA clone 139187 3'.											
		H.sapiens mRNA for uridine phosphorylase.											
		yn54c02.s1 Homo sapiens cDNA clone 172226 3' simil											
		EST17149 Homo sapiens cDNA 5' end similar to None.											
		Human gene for alpha I globin.											
		Human jun-B mRNA for JUN-B protein.											
		yj90e08.s1 Homo sapiens cDNA clone 156038 3'.											
		yj67b10.s1 Homo sapiens cDNA clone 153787 3'.											
		yj72b03.s1 Homo sapiens cDNA clone 154253 3'.											
		Human Na <sup>+</sup> ,K <sup>+</sup> ATPase gene exons 1 - 3 (alpha III is											
		Unknown											
		H.sapiens HCG I mRNA.											
		Homo sapiens porin (por) mRNA, complete cds and tr											
		Human 78 kDa gastrin-binding protein mRNA, complet											
		Human BENE mRNA, partial cds.											
		Human semaphorin V mRNA, complete cds.											
		Human HepG2 3'-directed Mbol cDNA, clone s150.											
		Human TNF-alpha inducible responsive element mRNA,											
		H.sapiens EDDRI gene for receptor tyrosine kinase.											
		ye38a04.s1 Homo sapiens cDNA clone 119982 3'.											
		za25g05.s1 Homo sapiens cDNA clone 293624 3'.											
		zb86e03.s1 Soares senescent fibroblasts NbHSF Homo sapiens cDNA											
		clone 310492 3'.											
		H.sapiens EST sequence (007-X1-01) from skeletal m											
		zr21b10.s1 Stratagene NT2 neuronal precursor 937230 Homo sapiens											
		cDNA clone 664027 3'.											
		Human heparin binding protein (HBP17) mRNA											

51	CATGGGATTCCAGTT	H671052	11	0	4	3	2	WS2456	zc4Se09.r1 Soares senescent fibroblasts NbHSF Homo
----	-----------------	---------	----	---	---	---	---	--------	----------------------------------------------------

# Transcripts decreased in both colon primary tumors and colon cancer cell lines compared to normal colon (130 genes)

NC: Normal Colon

TU: Colon Primary Tumor

CL: Colon Cancer Cell Line

PT: Pancreatic Primary Tumor

PC: Pancreatic Cancer Cell Line

#	Tag Sequence	Tag Number	NC	TU	CL	PT	PC	Accession	Gene Name
1	CATGCTCCAGCTAC	H382109	803	191	304	136	663	X12882	Human mRNA for cytokeratin 8.
2	CATGCTAAGACTTCA	H460926	708	282	402	142	497	F15636	H.sapiens mitochondrial EST sequence (002T15)
3	CATGGCCCAAGTCA	H610997	705	58	2	2	1		Unknown
4	CATGACCTTGGCCA	H90022	512	348	93	43	235	F16940	H.sapiens mitochondrial EST sequence (009-T1-21) f
5	CATGACATTGGGTGA	H81583	504	92	4	0	0	M10050	Human liver fatty acid binding protein (FABP) mRNA
6	CATGGCGAAACCTG	H622680	486	108	27	30	13	S61953	c-erbB3-receptor tyrosine kinase (alternatively sp
7	CATGAGCCCTACAA	H153361	367	242	132	71	204	F15506	H.sapiens mitochondrial EST sequence (1-i-02) from
8	CATGGACCCCAAGATA	H545828	276	131	0	7	0	T39321	ya04c01.r2 Homo sapiens cDNA clone 60480 5'.
								H24673	y141a01.s1 Homo sapiens cDNA clone 160776 3'.
									HUMGS02706 Human colon 3'directed Mbol cDNA, HUMGS02706, clone cm1673.
								D25586	clone cm1673.
								T96160	ye09b02.s1 Homo sapiens cDNA clone 117195 3'.
9	CATGGCCGGGTGGGC	H617195	256	88	148	144	178	X64364	H.sapiens mRNA for M6 antigen.
10	CATGTTGGGGTTTCC	H1026814	202	75	84	235	369	M11146	Human ferritin H chain mRNA, complete cds.
11	CATGCTCCACCCGAA (or G)	H479577	201	120	0	11	3	L15203	Human secretory protein (P1.B) mRNA, complete cds.
12	CATGGCAGGCCTCA	H600670	196	68	6	32	19	X93036	H.sapiens mRNA for MAT8 protein.
13	CATGATCGTGGCGGG	H224923	194	24	97	40	39	H93844	yv07h09.r1 Homo sapiens cDNA clone 242081 5' similar to SP:A39484
14	CATGCAAGCATCCCC	H271574	190	99	101	30	139	F17001	A39484 ANDROGEN-WITHDRAWAL APOPTOSIS PROTEIN RVPI.
15	CATGGACATCAAGTC	H544012	189	33	76	57	219	Y00503	H.sapiens mitochondrial EST sequence (011-T1-13) f
									Human mRNA for keratin 19.
									zb05a1.1.r1 Soares fetal lung NbHL19W Homo sapiens cDNA clone
									301148 5' similar to gb:V00567 BETA-2-MICROGLOBULIN
									PRECURSOR (HUMAN);
16	CATGGTTGTGTTAA	H782013	178	110	14	340	139	W16632	301148 5' similar to gb:V00567 BETA-2-MICROGLOBULIN
									zo31h04.s1 Striatogene colon (#937204) Homo sapiens cDNA clone
								AA143804	588535 3'



[illegible]



										AA303091	EST12940 Uterus tumor 1 Homo sapiens cDNA 3' end za52d02.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone 296163 5'.
63	CATGGCAGCTCCTGT	H599903	43	8	17	24	13			W02429	Yx44c11.s1 Homo sapiens cDNA clone 264596 3'.
										N20325	yz13c12.s1 Homo sapiens cDNA clone 282934 3'.
										N45127	zb38c11.s1 Soares parathyroid tumor NbHPA Homo sapiens cDNA clone 305876 3'.
										N90407	Human wild-type p53 activated fragment-1 (WAF1) mR
64	CATGTGTCCTGGTTC	H972720	43	12	14	25	5			U03106	zc11f01.s1 Soares parathyroid tumor NbHPA Homo sapiens cDNA clone 322009 3'.
65	CATGACAAACCCCA	H65878	42	16	7	12	11			W37827	gb1W15332 W15332 zc16d10.s1 Soares parathyroid tumor NbHPA cDNA clone 322483 3'.
										W15332	Homo sapiens zc04g10.s1 Soares parathyroid tumor NbHPA Homo sapiens cDNA clone 321378 3'.
										W32410	clone 321378 3'.
										N32312	yw82c01.s1 Homo sapiens cDNA clone 258720 3'.
										U51478	Human sodium/potassium-transporting ATPase beta-3
										Unknown	
66	CATGTAGGATGGGG	H828331	41	6	11	6	9				zp44f11.s1 Stratagene muscle 937209 Homo sapiens cDNA clone 612333 3' similar to contains Alu repetitive element;
67	CATGACTGTGGCGC	H126619	41	7	1	4	35			AA180815	yh87e04.s1 Homo sapiens cDNA clone 136734 3' similar to contains Alu repetitive element;
68	CATGGTAGCAGGTGT	H730287	40	7	13	17	24			R34696	yh87e04.s1 Homo sapiens cDNA clone 136734 3' similar to contains Alu repetitive element;
										R34696	repetitive element;
										AA194497	zq06e03.s1 Stratagene muscle 937209 Homo sapiens cDNA clone 628924 3' similar to contains Alu repetitive element
										AA194497	hbc760 Homo sapiens cDNA clone hbc760 3'end similar to nonspecific crossreacting antigen.
69	CATGAATCACAATA	H53508	40	12	0	3	0			T11144	zl67e01.s1 Stratagene colon (#937204) Homo sapiens cDNA clone 509688 3' similar to TR:G189087
										AA058357	zo31e02.s1 Stratagene colon (#937204) Homo sapiens cDNA clone 588506 3'.
										C05803	similar to none
										AA143765	zo31e02.s1 Stratagene colon (#937204) Homo sapiens cDNA clone 588506 3'.
70	CATGAGGATGGTCCC	H167606	40	11	4	4	5			AA179299	zp45b09.s1 Stratagene HeLa cell s3 937216 Homo sapiens cDNA clone 612377 3'.

71	CATGCCAAAGCTATA	H328308	38	11	6	2	18	M35252	Human CO-029.
72	CATGCGGAGTCGGG	H434907	38	8	6	0	0	R87448	ym89c10.s1 Homo sapiens cDNA clone 166098 3'.
73	CATGCCCGTGGAGAG	H618121	38	9	5	17	26	X79882	H.sapiens lrp mRNA.
74	CATGCCCCCGAAGCC	H349706	37	6	0	0	0		Unknown
75	CATGATTCAAGATG	H259108	37	1	0	0	0	J03037	Human carbonic anhydrase II mRNA, complete cds.
76	CATGGCCAGTGGCT	H611050	37	3	0	2	10		Unknown
77	CATGATGGTGGGGA	H241323	36	2	6	25	2	M92843	H.sapiens zinc finger transcriptional regulator mRNA
78	CATGCTGCCCCCT	H386390	35	12	7	7	5	X60188	Human ERK1 mRNA for protein serine/threonine kinase
79	CTAGTGGAAGTGAA	H950457	34	1	1	12	0	V01512	Human cellular oncogene c-fos (complete sequence).
80	CATGGTCATCACCAC	H740629	34	0	0	0	0	U34279	Human uroguanylin mRNA, complete cds.
81	CATGCTTATGGTCCC	H511670	34	1	0	3	1	AA287021	zs57c03.s1 Soares NbHTGBC Homo sapiens cDNA clone 701572 3'.
82	CATGCTGGGCCCTCTG	H502136	34	3	4	11	5	T55226	yb47a01.s1 Homo sapiens cDNA clone 74280 3' containing L1 repetitive element
								R37446	yf56e10.s1 Homo sapiens cDNA clone 26129 3' similar to gb:X07173 INTER-ALPHA-TRYPsin INHIBITOR COMPLEX COMPONENT II
								AA406180	zu65c08.s1 Soares testis NHT Homo sapiens cDNA clone 742862 3'.
83	CATGGCCCAAGGGCCC	H610982	33	3	0	0	2	R09752	Unknown
84	CATGTTTTTACTGAT	H1047673	33	7	0	4	2	R81530	yj02b10.r1 Homo sapiens cDNA clone 147547 5'.
								T32348	EST47211 Homo sapiens cDNA 3' end similar to None..
								zd17g02.s1	Soares fetal heart NbHH19W Homo sapiens cDNA clone
								W57810	340946 3'.
								zt47e12.s1	Soares ovary tumor NbHOT Homo sapiens cDNA clone
								AA398527	725518 3'.
85	CATGCCTGCTTGTCG	H387054	32	2	1	6	32	X63187	H.sapiens HE4 mRNA for extracellular proteinase inhibitor homologue
86	CATGACCTGGGGAGG	H96931	32	6	4	8	6		Unknown
87	CATGCCTTCAAATCA	H390158	31	1	0	0	0	R46266	y852g07.s1 Homo sapiens cDNA clone 36232 3' similar to gb:M33987 CARBONIC ANHYDRASE I
88	CATGTCGGAGCTGTT	H893564	30	1	4	7	1	H98618	yx12a06.s1 Homo sapiens cDNA clone 261490 3'.
								zo97h01.s1	Stratagene ovarian cancer (#937219) Homo sapiens cDNA clone 594865 3'.
								AA171705	
								H99212	yx15g08.s1 Homo sapiens cDNA clone 261854 3'.



[illegible]

99	CATGTACCTCTGATT	H810468	27	5	7	11	12	X65614	H.sapiens mRNA for calcium-binding protein S100P.
100	CATGATGATGGCACC	H233106	26	0	2	0	2		
101	CATGTTCTGTAGCCC	H1014566	25	5	0	4	0		embJZ6988  HISSECA3M H.sapiens mRNA for adenosine triphosphatase, calcium
102	CATGCCCTGTCTGCCA	H388582	24	1	2	1	3	T99568	ye65c02.r1 Homo sapiens cDNA clone 122594 5'.
								T87539	yd89D9.s1 Homo sapiens cDNA clone 115433 3'.
103	CATGTATGATGAGCA	H844682	23	4	0	1	0		gb AA347726 JAA347726 EST54132 Fetal heart II Homo sapiens cDNA 5' end similar to transmembrane secretory component
104	CATGCTGGCAAGGT	H500747	23	0	0	0	0		
105	CATGCTTGATCCCA	H517078	23	4	4	17	7	L42379	Homo sapiens bone-derived growth factor (BPGF-1) m
106	CATGCTTGACATACC	H516402	22	0	0	7	2	X68277	H.sapiens CL 100 mRNA for protein tyrosine phosphatase hydrolase
107	CATGGCTGGCACATT	H649492	22	5	0	0	0	M82962	Human N-benzoyl-L-tyrosyl-p-amino-benzoic acid alpha subunit (PPH alpha) mRNA, complete cds
108	CATGCTCTGAATTATG	H909556	21	1	1	1	1	X16354	Human mRNA for transmembrane carcinoembryonic antigen (CEA)
109	CATGGGAAGACCACT	H657554	21	1	1	3	3	X74570	H.sapiens mRNA for Gal-beta(1-3/1-4)GlcNAc alpha-2,3-sialyltransferase
110	CATGGCTCTTCCCCA	H646998	20	2	0	1	0	R87768	yo45d01.s1 Homo sapiens cDNA clone 180865 3' similar to contains PTR5 repetitive element
								R85880	yo36g07.s1 Homo sapiens cDNA clone 180060 3' similar to contains PTR5 repetitive element
111	CATGAAATCTGGCAC	H14245	20	2	0	4	3	L20826	Human I-plastin mRNA, complete cds.
112	CATGTAAATTGCATT	H802708	19	2	0	1	7	Z50751	HSB4BMR H.sapiens mRNA for B4B
								U77085	Human epithelial membrane protein (CL-20) mRNA, complete cds
								Y07909	HSPAPR H.sapiens mRNA for Progression Associated Protein
113	CATGGTGGGGCGCC	H764570	18	1	1	8	2	R48329	yj64g10.r1 Homo sapiens cDNA clone 153570 5'.
114	CATGTTATGGTGTGA	H998127	17	0	0	1	0	T27534	EST10a24 Clontech adult human fat cell library HL1108A Homo sapiens cDNA clone 10a24.
115	CATGGGAGAAACAGC	H663571	17	1	2	4	0	T86124	yd84b04.s1 Homo sapiens cDNA clone 114895 3'.
								zo15g05.s1	Stratagene colon (#937204) Homo sapiens cDNA clone 587000 3'.
								R49945	yj58g11.s1 Homo sapiens cDNA clone 152996 3'.
								T57044	ya84b01.s1 Homo sapiens cDNA clone 68401 3'.
116	CATGCCCAACACCAGC	H328787	17	1	0	0	0		
117	CATGAGGTGACTGGG	H178299	17	0	0	0	0		
118	CATGGCCATCCTCCA	H609634	16	0	0	0	0		gb R73013 R73013 yj94a09.r1 Homo sapiens cDNA clone 156376 5'.

		H1039799	15	1	0	4	4	M69013	Human guanine nucleotide-binding regulatory protein
119	CATGTTTCTCGTCGC			1	0	4	0		Unknown
120	CATGTCAGAGCGCTG	H860776	15	1	1	1	0		yy72h06.s1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone 248315 3' similar to contains element PTR7 repetitive element
121	CATGTTCCCGCTTCC	H1006014	14	1	0	0	2	N58523	Unknown
122	CATGTACGGTGTGGG	H814011	14	1	0	0	0		Unknown
123	CATGCTCAGAACTTG	H477216	14	0	1	4	13		Unknown
124	CATGGGACTAAATGA	H662543	13	1	0	1	0	M29540	Human carcinoembryonic antigen mRNA (CEA), complete cds.
125	CATGGCTTGGGGATT	H653988	12	0	0	0	1	D25786	HUMGS04154 Human colon 3 directed Mbol cDNA, HUMGS04154, clone cm0215.
									yc36e02.r1 Homo sapiens cDNA clone 82778 5' similar to gb:L07765 LIVER CARBOXYLESTERASE PRECURSOR
								T73613	Unknown
126	CATGACCCCAACTGCC	H86138	12	0	0	0	1		gbT95615JT95615 ye40e03.s1 Homo sapiens cDNA clone 120220 3'
127	CATGCTGAACCTCCC	H491894	12	0	0	2	2		zf19b11.s1 Stratagene NT2 neuronal precursor 937230 Homo sapiens cDNA clone 663837 3'
128	CATGCAAGAGTTTCT	H271102	11	0	0	2	0	AA226797	zq97h01.s1 Stratagene NT2 neuronal precursor 937230 Homo sapiens cDNA clone 649969 3'
								AA218730	yp57f10.r1 Homo sapiens cDNA clone 191563 5' similar to gb:M90657 TUMOR-ASSOCIATED ANTIGEN L6 (HUMAN);.
129	CATGOTCCGAGTGCA	H743610	11	0	0	8	5	H38178	Unknown
130	CATGTTTGTTTCAC	H1043445	11	0	0	0	0		

# Transcripts decreased in only colon cancer cell lines compared to normal colon (78 genes)

NC: Normal Colon

TU: Colon Primary Tumor

CL: Colon Cancer Cell Line

PT: Pancreatic Primary Tumor

PC: Pancreatic Cancer Cell Line

#	Tag sequence	Tag Number	NC	TU	CL	PT	PC	Accession	Gene Name
1	CATGCACCTAATTGG	H285759	612	755	411	161	333	F15516	H.sapiens mitochondrial EST sequence (1-t-12)
2	CATGATTGAGAAAGC	H260227	603	566	158	249	173	F12396	H. sapiens partial cDNA sequence; clone c-39e04.
3	CATGTGATTTCACCT	H933704	452	595	235	80	314	L08441	Human autonomously replicating sequence (ARS) mRNA
4	CATGTTTCATACACCT	H1002566	444	357	114	64	191	F15553	H.sapiens mitochondrial EST sequence (001T14)
5	CATGCCACTGCACTC	H335432	385	402	223	278	132	X51525	Human cortex mRNA containing an Alu repetitive element
6	CATGACTAACACCT	H114966	369	446	171	76	161	F16402	H.sapiens mitochondrial EST sequence (141-20)
7	CATGCACTACTCACC	H291282	293	527	78	14	83	U09500	Human mitochondrion cytochrome b gene, partial cds
8	CATGAAACATTCCTC	H1272	200	169	98	17	223	F15744	H.sapiens mitochondrial EST sequence (101-03)
9	CATGCTCATAAGGAA	H478249	184	127	70	21	75	F15511	H.sapiens mitochondrial EST sequence (1-t-07)
10	CATGTCGAAGCCGCC	H885334	147	183	94	49	57	F18587	H.sapiens mitochondrial EST sequence (022T19)
11	CATGACGCAGGGAGA	H103075	145	160	91	69	47	H03983	yj47a08.s1 Homo sapiens cDNA clone [51862 3'.
12	CATGTTGGCCAGGCT	H1025322	124	194	63	111	51	X74301	H.sapiens mRNA for MHC class II transactivator.
13	CATGTTGGTGAAGGA	H1027595	98	106	17	183	107	M17733	Human thymosin beta-4 mRNA, complete cds.
14	CATGATCACGCCCTC	H214616	97	186	17	41	49	U46913	Human EST overexpressed in pancreatic cancer (xs31)
15	CATGTGCCCTGCACCA	H941638	67	48	25	75	34	X05607	Human mRNA for cysteine proteinase inhibitor precursor
16	CATGAGACCCACAAC	H136465	64	121	28	24	15	D34113	Human fetal brain cDNA 5'-end GEN-129B05.
17	CATGAGTTTGTAGT	H196339	60	33	17	13	15	X14758	Human mRNA for adenocarcinoma-associated antigen
18	CATGGGAACAACAG	H656389	56	41	4	31	3	L33930	Homo sapiens CD24 signal transducer mRNA
19	CATGTGGTGTATGCA	H965434	53	271	6	30	5	D50954	Human fetal brain cDNA 3'-end GEN-002A10.
20	CATGGAAATACAGTT	H527436	49	35	10	100	36	M11233	Human calthepsin D mRNA, complete cds.
21	CATGGTGGCTCACGC	H763719	49	37	21	27	15	U25801	Human Tax I binding protein mRNA, partial cds.
22	CATGGTGGTGACAC	H765509	45	26	18	23	15	U31215	Human metabotropic glutamate receptor 1 alpha
23	CATGGGGTTGGCTTG	H704160	44	56	2	6	1	S79597	(RNA)Ser(UNC) [human, muscle, MERRF/MELAS overlap s
24	CATGGTGGCGGTGC	H763567	42	32	15	20	5	T48809	yb05c03.r1 Homo sapiens cDNA clone 70276 5' contai
25	CATGTAGACTAGCAA	H821029	39	23	1	23	10	M69023	Human globin gene.

26	CATGGCTAGGTTTAT	H641789	38	144	13	25	13	DS1017	Human fetal brain cDNA 3'-end GEN-007C04
27	CATGGGCTTAGGGA	H687915	37	372	6	29	11	W15552	z991h1.s1 Soares parathyroid tumor NbHPA Homo sap
28	CATGGGGGTCAGGG	H699691	37	170	11	16	9	F16326	H.sapiens mitochondrial EST sequence (132-20) from skeletal muscle
29	CATGATTTTCTAAAA	H261569	33	13	11	8	2	AA315049	EST186995 HCC cell line (metastasis to liver in mouse) II Homo sapiens cDNA 5' end
30	CATGCACTTGCCCT	H294488	33	18	11	17	36	F01150	H.sapiens partial cDNA sequence; clone A6A03; ver
31	CATGCCCTGCTGCAGG	H386963	32	13	0	6	2	N29971	yw53h01.s1 Homo sapiens cDNA clone 255985 3'
32	CATGAGAACCTTCCA	H132598	32	14	3	16	12	K02883	Human MHC class I HLA-A2 gene, complete cds.
33	CATGCTCTGCCCTC	H489822	32	32	7	20	5	R09140	y25f12.s1 Homo sapiens cDNA clone 127919 3'
								R76005	y122c10.s1 Homo sapiens cDNA clone 158994 3'
								T33596	EST58371 Homo sapiens cDNA 3' end similar to None..
34	CATGGCCATCCCTT	H609624	29	73	7	14	16	F16449	H.sapiens mitochondrial EST sequence (129-09)
35	CATGGCCACGGGCC	H610922	28	9	1	1	7	AA292959	z54P10.s1 Soares ovary tumor NbHOT Homo sapiens cDNA clone 726187 3'
36	CATGTGGCCGCTGTC	H956860	26	8	1	1	2	AA292466	z31c11.r1 Soares ovary tumor NbHOT Homo sapiens cDNA clone 723956 5' similar to TR:G205858 G205858 RAT ORF
								N92384	zb62d07.s1 Soares fetal lung NbHL19W Homo sapiens cDNA clone 308173 3' similar to PIR:A39484 A39484 androgen-withdrawal apoptosis protein RVP1, prostate - rat
								N80203	zb19c06.s1 Homo sapiens cDNA clone 302506 3' similar to PIR:A39484 A39484 androgen-withdrawal apoptosis protein RVP1, prostate - rat;
									zk39d06.s1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone 485195 3' similar to PIR:A39484 A39484 androgen-withdrawal apoptosis protein RVP1
37	CATGAGGGTGTTTC	H175872	26	218	7	20	10	AA039323	Human partial cDNA sequence with CCA repeat region
38	CATGCCCTGGGAAGTG	H387596	25	10	0	45	17	U21468	Human epistatin variant A mRNA, 3' end.
39	CATGAGTCTGCTGGA	H188027	24	9	1	0	0	M34088	Unknown
40	CATGCCCGCCTCTTC	H353760	24	11	2	3	4	T10098	seq816 Homo sapiens cDNA clone b4HB3MA-COT8-HAP-F1
41	CATGAAAAGAGTGGT	H2235	22	9	2	0	7	X83228	H.sapiens mRNA for LI-cadherin.
42	CATGGCCACGTGGAG	H607977	21	7	1	2	2	L27415	Homo sapiens huntingtin (HD) gene, exon 66.
								C00470	dbj C00470 C00470 HUMGS0007620, Human Gene Signature, 3'-directed cDNA sequence.
43	CATGAGGATGTGGG	H167659	21	5	4	1	3	N63531	y62g08.s1 Homo sapiens cDNA clone 278174 3'

[illegible]

62	CATGGGGCTAGTCC	H695406	14	4	0	1	0	M25629	Human kallikrein mRNA, complete cds, clone clone p
63	CATGCCCGGCTCCTC	H134776	14	7	1	5	2	H18836	ym45d10.s1 Homo sapiens cDNA clone 51262 3'
								AA026974	zK01e10.s1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone 469290 3'
								AA405031	zu12c12.r1 Soares testis NHT Homo sapiens cDNA clone 731638 5' similar to gb:M61900 Human prostaglandin D synthase gene, complete cds. (HUMAN); gb U66894 HSU66894 Human epithelium-restricted Ets protein ESX mRNA, Human epithelial-specific transcription factor ESE-1b (ESE-1) mRNA, complete cds
64	CATGAGGTACTACTA	H1176584	13	9	0	9	8	U66894	Human epithelial-specific transcription factor ESE-1b (ESE-1) mRNA, complete cds
								U73843	Human colon 3'directed Mbol cDNA, HUMGS06772
65	CATGCAAAATAAATTA	H26232	13	3	0	1	0	D25996	Unknown
66	CATGCTGTAAAAAAA	H503809	13	6	0	1	1		ze88g07.s1 Soares fetal heart NbHH19W Homo sapiens cDNA clone 366108 3'
67	CATGGTTCAATCCCT	H774358	13	3	0	2	0	AA071520	za90h10.s1 Soares fetal lung NbHL19W Homo sapiens cDNA clone 299875 3'
								N90742	zn52h06.s1 Stratagene muscle 937209 Homo sapiens cDNA clone 561851 3'
								AA086292	Human HepG2 3'-directed Mbol cDNA, clone a-35.
68	CATGAATAAAGCCTT	H49304	12	4	0	0	0	D11499	IB2474 Homo sapiens cDNA 3'end.
69	CATGGGAAGGTTTAC	H658173	12	2	0	1	0	T16031	yc82e01.r1 Homo sapiens cDNA clone 22306 5'
70	CATGGGATGGCTTAT	H670333	12	1	0	6	1	T74426	za61h02.s1 Homo sapiens cDNA clone 297075 3'
71	CATGGGTGGCCCGGG	H715099	12	2	0	3	2	N73771	zh75f08.s1 Soares fetal liver spleen INFLS S1 Homo sapiens cDNA clone 417927 3'
								W90388	H. sapiens partial cDNA sequence, clone c-29h08.
								F03786	Human 11 beta-hydroxysteroid dehydrogenase type II
72	CATGTACTGTACTTC	H817952	12	2	0	0	0	U14631	ya31a06.s5 Homo sapiens cDNA clone 62194 3' contains Alu repetitive element.
73	CATGCCCTTGCACTC	H360008	11	6	0	3	3	T41121	Unknown
74	CATGCGGTGGGACCA	H440966	11	4	0	2	0		Unknown
75	CATGGCCCCCAACCA	H611590	11	2	0	0	0		Unknown
76	CATGGCCGGCGCTC	H616862	11	2	0	0	0	Z58486	Unknown
77	CATGGGAGGGCGCTCA	H666014	11	1	0	0	0		Unknown

78	CATGTCCCCCGTTACA	H874226	11	11	0	0	0	0	W68073	zd42c12.s1 Soares fetal heart NbHH19W Homo sapiens cDNA clone 343318 3' similar to contains Alu repetitive element;
----	------------------	---------	----	----	---	---	---	---	--------	------------------------------------------------------------------------------------------------------------------------



Table 4 - Transcripts increased in pancreas.cancer -  
SAGE Tags elevated only in Pancreatic Tumor

NC Normal Colon  
Tu Colon Tumor  
CC Colon Cancer Cell Line  
PT Pancreatic Tumor  
PC Pancreatic Cell Line

Tag Sequence	Tag Number	NC	Tu	CC	PT	PC	Examples	Accession	Gene Name
1 CATGAAGCAACCA	H9222	0	6	1	3	11	Examples	R38305	yh95b04.s1 Homo sapiens cDNA clone 137455 3'
								AA126719	zk95b03.s1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone 490541 3'
								AA044296	zk51c03.s1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone 486340 3'
								AA131586	zl33c08.s1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone 503726 3'
								AA157983	zo71h12.s1 Stratagene pancreas (#937208) Homo sapiens cDNA clone 592391 3'
2 CATGAAGCAGTTTA	H9408	1	5	2	21	3	Examples	AA292929	zi54e04.s1 Soares ovary tumor NbHOT Homo sapiens cDNA clone 726174 3'
								AA159306	zo78c07.s1 Stratagene pancreas (#937208) Homo sapiens cDNA clone 154129 3'
								R54012	yj70h01.s1 Homo sapiens cDNA clone 79335 3'
								T62936	yb99j08.s1 Homo sapiens cDNA clone 79335 3'
								X52426	H. sapiens mRNA for cytokeratin 13
3 CATGAAGCGGGGCT	H9898	0	0	0	0	13	Examples	X51698	H. sapiens spasmolytic polypeptide (SP) mRNA.
4 CATGAATCCTGGGT	H13803	0	1	1	16	2	Examples	N70419	za61d12.s1 Homo sapiens cDNA clone 297047 3'
5 CATGAATGGACAAC	H14865	0	0	1	0	13	Examples	AA411599	zv16g01.r1 Soares NbHMPu S1 Homo sapiens cDNA clone 753840 5'
								AA410508	zv16g01.s1 Soares NbHMPu S1 Homo sapiens cDNA clone 753840 3'
								AA115723	zl86g12.s1 Stratagene colon (#937204) Homo sapiens cDNA clone 511538 3'
6 CATGAACCAGTTGT	H21247	1	1	6	8	13	Examples	AA132875	zo19e04.s1 Stratagene colon (#937204) Homo sapiens cDNA clone 587358 3'
								AA147677	zo44a06.s1 Stratagene endothelial cell 937223 Homo sapiens cDNA clone 589714 3'

[illegible]

										AA279290	zs84a06.s1 Soares NbHTGBC Homo sapiens cDNA clone 704146 3'
										AA046253	z112a02.s1 Soares fetal heart NbIH19W Homo sapiens cDNA clone 376682 3'
15	CATGACAACTCAATA	H67396	2	7	16	37			Examples	Z58016	H.sapiens CpG DNA, clone 36c7.
										AA151668	zo29c02.s1 Stratagene colon (#937204) Homo sapiens cDNA clone 588290 3' similar to SW:B13_MOUSE P28662 BRAIN PROTEIN I3
										W02958	za07e06.r1 Soares melanocyte 2NbHM Homo sapiens cDNA clone 291874 5'
										AA1556464	zo70e05.s1 Stratagene pancreas (#937208) Homo sapiens cDNA clone 592256 3'
16	CATGACACCCCTGTGC	H71151	0	1	0	2	14		Examples	AA025673	ze90h09.s1 Soares fetal heart NbHFH19W Homo sapiens cDNA clone 366305 3'
										N70895	za89h12.s1 Homo sapiens cDNA clone 299783 3'
										X02491	Human interferon-inducible mRNA (cDNA 9-27): membrane
17	CATGACCATTTGGATT	H85924	0	8	5	13	4		Examples	J04164	Human interferon-inducible protein 9-27 mRNA
										X84958	H.sapiens mRNA for interferon-induced 17kDa membra
										X56841	H.sapiens HLA-E gene.
18	CATGACCCTTTAACA	H90050	1	4	2	13	7		Examples	X64879	H.sapiens mRNA for HLA-E heavy chain (exons 4 - 7)
										M21186	Human neutrophil cytochrome b light chain p22A
19	CATGACCGCCGTGGT	H91579	49	22	45	70	94		Examples	M61107	Human p22-phox (CYBA) gene, exons 3 and 4
										D00244	Human Pro-urokinase gene.
20	CATGACCTGTGACCA	H97158	0	3	0	28	17		Examples	K02286	Human urokinase gene, 3' end
										M15476	Human pro-urokinase mRNA, complete cds
										X02419	Human uPA gene for urokinase-plasminogen activator
										L08835	Human myotonic dystrophy kinase (DM kinase) gene
21	CATGACGCCCTGCTC	H103912	0	1	0	11	2		Examples	M87313	Homo sapiens myotonic protein kinase (DM) mRNA
										H44451	yo75f06.s1 Homo sapiens cDNA clone 183779 3'
22	CATGACGTGGTGATG	H113380	2	4	4	5	20		Examples		zo42i07.s1 Stratagene endothelial cell 937223 Homo sapiens cDNA clone 589573 3' similar to SW:L10K_RAT Q05310 LEYDIG CELL TUMOR 10
										AA157329	KD PROTEIN
											zc32g06.s1 Soares senescent fibroblasts NbHSF Homo sapiens cDNA clone 324058 3' similar to SW:L10K_RAT Q05310 LEYDIG CELL TUMOR 10
										W46455	KD PROTEIN

23	CATGACTCAGCCCGG	H119383	0	0	3	21	3	Examples	M92357	Homo sapiens B94 protein mRNA, complete cds.
24	CATGACTGAGGAAG	H123521	0	0	0	53	22	Examples	X64875	H. sapiens mRNA for insulin-like growth factor binding protein 3 Human growth hormone-dependent insulin-like growth factor binding protein 3
									M31159	Human insulin-like growth factor-binding protein-3
									M35878	insulin-like growth factor binding protein 3 (3' region)
									S56205	Human extracellular matrix protein 1 (ECM1) mRNA
25	CATGACTGCCCGCTG	H124264	1	0	0	22	9	Examples	U65932	Human extracellular matrix protein 1 (ECM1) gene, exon 9
									U65937	z003709.s1 Stralagene colon (#937204) Homo sapiens cDNA clone 566633
26	CATGACTGTATTTC	H126208	3	4	9	2	22	Examples	AA148916	3'
									AA129137	z012a11.s1 Stralagene colon (#937204) Homo sapiens cDNA clone 586652
									AA115437	3'
									AA126967	z185g09.s1 Stralagene colon (#937204) Homo sapiens cDNA clone 511456
									AA126967	3'
27	CATGAGCTGCAGC	H149395	1	2	6	3	16	Examples	R24613	z187e07.s1 Stralagene colon (#937204) Homo sapiens cDNA clone 511620
28	CATGAGCAGGAGCGT	H150055	1	0	0	0	15	Examples	H43243	yh36c03.r1 Homo sapiens cDNA clone 131812
29	CATGAGCTGTATTCT	H162622	0	2	0	1	11	Examples	X54942	yp05e05.r1 Homo sapiens cDNA clone 186560 5'
										H.sapiens cks2 mRNA for Cks1 protein homologue
										zK50g07.s1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone
										486300 3'
30	CATGAGGATGACCCC	H167446	1	7	12	10	13	Examples	AA044081	zK50g07.r1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone
										486300 5' similar to PIR: A40533 A40533 cAMP-dependent protein kinase major membrane substrate
									AA044211	Class A, Human mRNA for thrombospondin.
31	CATGAGGTCTTCAAT	H178129	4	2	0	60	2	Examples	X14787	yh64n1.s1 Homo sapiens cDNA clone 134541 3'
32	CATGAGGTGCGGGG	H178603	0	2	2	1	11	Examples	R27738	yj22n2.s1 Homo sapiens cDNA clone 149519 3' similar to SP.ZK637.5
									H00276	CE00436 ARSA
										zm19d07.s1 Stralagene pancreas (#937208) Homo sapiens cDNA clone
33	CATGAGTATCTGGGA	H183787	3	3	1	15	73	Examples	AA076235	526093 3'
									H13159	yj16g04.s1 Homo sapiens cDNA clone 148902 3'
										zo71e11.s1 Stralagene pancreas (#937208) Homo sapiens cDNA clone
									AA146632	592364 3'
34	CATGATCTTAAAT	H204740	1	0	3	18	9	Examples	X80062	H.sapiens SA mRNA
									U01691	Human annexin V (ANX5) gene

							X12454	Human mRNA for vascular anticoagulant
							M18366	Human placental anticoagulant protein (PAP) mRNA
							M21731	Human lipocortin-V mRNA, complete cds
							J03745	Human endonexin II mRNA, complete cds
								GAMMA-INTERFERON-INDUCIBLE PROTEIN IP-30 PRECURSOR (HUMAN)
		H213518	2	1	5	25	I Examples J03909	EST97384 Thymus II Homo sapiens cDNA 3' end similar to interferon, gamma transducer 1
	CATGATCAACAATCC						aa383911	Human ribosomal protein L9 mRNA
		H213679	12	9	25	156	E Examples U09553	Human ribosomal protein L9 mRNA, complete cds
	CATGATCAAGGGTGT						U21138	Human mRNA for human homologue of rat ribosomal protein zm03a05.s1 Striatogenic corneal stroma (#937222) Homo sapiens cDNA clone 513008 3'
							D14531	RNA polymerase II transcription factor SIII p18 subunit mRNA
		H213751	0	2	8	3	10	H.sapiens CpG DNA, clone 13a10, reverse read cpg 1
	CATGATCAAACTTCGA						E Examples AA063259	
		H219750	16	7	14	12	40	
	CATGATCGGCGCCA						E Examples L42856	
		H229502	1	0	0	17	4	
	CATGATGAAACTTCG						E Examples Z59242	
		H235531	2	3	12	3	22	
	CATGATGCGRAAGGC						E Examples Z25820	H.sapiens mRNA for mitochondrial dodecenoyl-CoA dehydrogenase
							L24774	Homo sapiens delta3, delta2-CoA-isomerase mRNA
		H243676	0	0	1	0	14	40S RUBOSOMAL PROTEIN S3A (HUMAN)
	CATGATGCTCTCGTT						E Examples M84711	Human insulin-like growth factor binding protein 4
		H243710	1	2	1	14	2	Human insulin-like growth factor binding protein-4 (IGFBP4) gene, promoter and complete cds
	CATGATGCTTTCT						U20982	H.sapiens mtsl gene.
		H244487	0	4	5	44	94	Human CAPL protein mRNA, complete cds
	CATGATGTGTAAACGA						E Examples Z33457	yx70b09.s1 Homo sapiens cDNA clone 267065 3' similar to gb:L12350 THROMBOSPONDIN 2 PRECURSOR (HUMAN)
		H270083	0	1	2	10	1	
	CATCCAACCTTAAGC						E Examples N23207	z125e11.s1 Soares ovary tumor NbHOT Homo sapiens cDNA clone 714188 3' similar to gb:M33680 CD81 ANTIGEN (HUMAN)
		H286424	0	4	2	10	1	
	CATGCACCTGTCCTT						E Examples AA285023	CD81 antigen
							M33680	Neurosin
		H291889	0	0	2	3	19	protease M
	CATGCACTCAATAAA						E Examples D78203	
							U62801	

47	CATGCAGCCTGGGC	H300971	0	0	0	0	0	0	10	Examples	AA149942	z668d04.s1 Stratagene pancreas (#937208) Homo sapiens cDNA clone 592039 3' similar to TR:E218488 E218488 TRYPTASE
48	CATGCAGCGGCCCT	H301462	4	11	12	10	21			Examples	AA187553 M16937	zp66b09.r1 Stratagene endothelial cell 937223 Homo sapiens cDNA clone 625145 5' similar to gb:M16937 HOMEOBOX PROTEIN HOX-B7 (HUMAN); contains element MER22 repetitive element Homeobox protein HOX-B7
49	CATGCAGGTTGTCT	H307126	0	0	4	0	10			No Match		Human ribosomal protein S10 mRNA
50	CATGCAGTCTCTCAA	H309109	2	6	6	2	17			Examples	U14972	Human leukotriene A4 hydrolase gene
51	CATGCATCCCGTGAC	H316857	0	3	3	3	13			Examples	U27293	Human leukotriene A-4 hydrolase mRNA, complete cds
											J03459	Human leukotriene A-4 hydrolase mRNA, complete cds
											J02959	Human leukotriene A-4 hydrolase mRNA, complete cds
52	CATGCATTCCTCCTT	H325080	0	2	5	13	3			Examples	X82434	H.sapiens mRNA for emerlin
53	CATGCCACCCACC	H33138	3	7	17	18	2			Examples	M88338	Human serum constituent protein (MSE55) mRNA
54	CATGCCAGTGGCCCG	H339606	23	11	37	22	56			Examples	U14971	Human ribosomal protein S9 mRNA
55	CATGCCATTTCTGG	H344031	0	2	6	1	10			Examples	L01697	Homo sapiens alpha-1 type XV collagen mRNA
56	CATGCCCAAGCTAGC	H344691	19	8	8	18	44			Examples	X54079	Human mRNA for heat shock protein HSP27.
											Z23090	H.sapiens mRNA for 28 kDa heat shock protein
											X16477	Human mRNA fragment for estrogen-regulated 24k protein
											S74571	estrogen receptor-related protein=27-kda heat shock protein
57	CATGCCCATCCGAAA	H347489	20	15	43	19	61			Examples	X69392	H.sapiens mRNA for ribosomal protein L26.
											L07287	Human ribosomal protein L26 (RPL26) gene
58	CATGCCCCCTGCAGA	H350099	0	1	6	14	25			Examples	U40434	Human mesothelin or CAK1 antigen precursor mRNA
											D49441	Human mRNA for pre-pro-megakaryocyte potentiating factor, complete cds.
59	CATGCCCGCATAGAT	H353481	0	0	0	8	11			Examples	U12819	Human p16-INK4 (p16) gene
											U38945	Human hypothetical 18.1 kDa protein (CDKN2A) mRNA
											S69804	MTS1= multiple tumor suppressor 1/cyclin-dependent kinase 4 inhibitor p16
											S69822	CDK41=cyclin-dependent kinase 4 inhibitor
											S78535	tumor suppressor gene, P16/MTS1/CDKN2=cell cycle negative regulator beta form
(60)	CATGCCCTCTGGGG	H357867	8	2	5	14	34			Examples	Z47319	H.sapiens mRNA for expressed sequence tag (clone 2167119)



									M11233	Human cathepsin D mRNA, complete cds y442f03.s1 Homo sapiens cDNA clone 110909 3' similar to SP-R151.9 CE00827
98		H57929	4	7	5	14	26	Examples	T90296	
		CATGGAATGATGAG							AA320942	EST73523 Adipose tissue, brown Homo sapiens cDNA 3' end zp64f07.s1 Stratagene endothelial cell 937223 Homo sapiens cDNA clone 624997 3'
99										
100		H533436	3	7	16	6	28	Examples	AA181811	z106c06.s1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone 491530 3' similar to WP.ZK652.2 CE00448
		CATGGAAGATGTGTG							AA148508	Human peripheral benzodiazepine receptor related mRNA Human peripheral benzodiazepine receptor (hpbs) mRNA
101										
102		H540621	6	3	10	9	28	Examples	L21950	
		CATGGATTTTATAA							M36035	
103										
104		H540673	1	2	10	3	17	No Match		Human microfibril-associated glycoprotein (MFAP2).
105		CATGGACAAAAAAA							U19718	H.sapiens epithelial tropomyosin (TM1) mRNA
106		H545152	0	1	0	11	2	Examples	M75165	Human fibroblast muscle-type tropomyosin mRNA
107		CATGGACCACCTTTA							M12125	Human tropomyosin-1 (TM-beta) mRNA, complete cds
108		H545430	0	3	0	20	18	Examples		Human cyclin mRNA
109		CATGGACCAGGCCCT							M74817	Homo sapiens FK-506 binding protein homologue zb37g02.s1 Soares parathyroid tumor NbHPA Homo sapiens cDNA clone 303810 3'
110										z106a10.s1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone 491514 3'
111		H546059	2	5	9	16	10	Examples	M74092	Human platelet-derived endothelial cell growth factor
112		CATGGACCCCAAGC								Human gamma-tubulin mRNA, Human mRNA (HA1753) for ORF
113		H546710	31	36	20	71	65	Examples	L37033	Human mRNA (HA1753) for ORF
114		CATGGACCTGCCCT								TMMP-1=metalloproteinase inhibitor EPA glycoprotein (erythroid-potentiating activity) tissue inhibitor of metalloproteinase 2
115										
116		H548062	0	1	0	13	1	Examples	N90046	
117		CATGGACCTACTCT							AA115048	
118										
119		H551315	3	4	5	32	3	Examples	M63193	
120		CATGGACGGCGCAGG								
121		H554876	1	4	3	0	14	Examples	M61764	
122		CATGGACTCTCTGTT								
123		H559615	0	0	0	2	10	Examples	D17793	
124		CATGGAGAGCTTTGC							S68252	
125		H560056	0	5	8	32	11	Examples	X02598	
126		CATGGAGAGTGCTG							X03124	
127										
128										
129		H561807	0	0	0	1	12	No Match		
130		CATGGAGCAGGATGA								
131										
132		H567486	1	1	0	4	13	Examples	AA214523	zr89c01.s1 Soares NbHTGBC Homo sapiens cDNA clone 682848 3' yw75d01.s1 Homo sapiens cDNA clone 258049 3'
133		CATGGAGGGAGTTC							N30324	H.sapiens mRNA for neurotensin receptor.
134										
135		H570787	0	0	2	1	10	Examples	X70070	yrr27a10.s1 Homo sapiens cDNA clone 206490 3'
136		CATGGAGTCCGAGC								
137		H572656	0	0	3	0	10	Examples	H57673	
138		CATGGAGTTATGTTG								





									M73239	Human (clone SF1) hepatocyte growth factor (HGF)
									M73240	Human (clone SF2) hepatocyte growth factor (HGF)
									X02920	Human mRNA for alpha 1-antitrypsin carboxyterminal, 0
109	CATGGGAAAAGTGGT	H655547	18	13	3	70	1	Examples	X01683	Human mRNA for alpha 1-antitrypsin
									V00496	Human messenger RNA for alpha-1-antitrypsin
									J00067	Human alpha-1 antitrypsin gene, 3' end
									Z12B01.s1	Soares pregnant uterus NbHPU Homo sapiens cDNA clone
									S02633 3'	
110	CATGGGAAGGAGGC	H658059	0	0	4	6	16	Examples	AA127040	zdr6f06.s1 Soares fetal heart NbHH19W Homo sapiens cDNA clone
									W81387	347555 3'
									H45477	yo7zh08.s1 Homo sapiens cDNA clone 183519 3'
									D26598	Human mRNA for proteasome subunit Hsc10-II, 0
111	CATGGGAGTCATTGT	H666943	6	5	6	10	32	Examples	N74310	za78c01.s1 Homo sapiens cDNA clone 298656 3'
									H92750	y192e01.s1 Homo sapiens cDNA clone 231768 3'
112	CATGGGAGTGCGT	H667367	0	0	1	1	10	Examples	T24084	seq2272 Homo sapiens cDNA clone ssb4HB3MA(extended-ft-6) 3'
									X17567	H sapiens RNA for snRNP protein B
113	CATGGGATTGCTGG	H671455	3	7	13	5	21	Examples	M34081	Human small nuclear ribonucleoprotein particle SmB
									M69054	Human insulin-like growth factor binding protein 6, 0
114	CATGGGCCCTCACC	H677330	0	0	2	9	22	Examples	M62402	Human insulin-like growth factor binding protein 6
									N74323	za78d08.s1 Homo sapiens cDNA clone 298671 3'
115	CATGGGCCCTCTGAG	H677753	0	1	4	7	14	Examples	H46766	yo18f08.s1 Homo sapiens cDNA clone 178311 3'
									H41102	yn88a08.s1 Homo sapiens cDNA clone 175478 3'
									AA074777	zm84b09.s1 Stratagene ovarian cancer (#937219) Homo sapiens cDNA clone 544601 3'
116	CATGGGCTGGTCTGG	H686815	0	1	3	13	22	Examples	AA062735	zm04a04.s1 Stratagene corneal stroma (#937222) Homo sapiens cDNA clone 513102 3'
									AA112905	zm63f12.s1 Stratagene fibroblast (#937212) Homo sapiens cDNA clone 530351 3'
117	CATGGGGAACAGAT	H688713	25	7	9	0	72	No Match		
118	CATGGGAGGGGTGG	H690863	2	3	1	16	2	No Match		
119	CATGGGAGGTAGCA	H690890	1	0	1	14	1	No Match		
120	CATGGGCATCTCTT	H693112	1	1	3	39	2	Examples	V00523	Human mRNA for histocompatibility antigen HLA-DR
									X00274	Human gene for HLA-DR alpha heavy chain a class II
									K01171	Human HLA-DR alpha-chain mRNA

											J00202	human hla-dr heavy chain gene; 3' flank
121	CATGGCTGGGAGAT	H715401	1	4	10	10	14		Examples	U18009		Human chromosome 17q21 mRNA clone LF113.
										T33413		EST57778 Homo sapiens cDNA 3' end similar to None
										T33339		EST57474 Homo sapiens cDNA 3' end similar to None
122	CATGGTACTGTAGCA	H728778	3	3	1	16	30		Examples	M59911		Human integrin alpha-3 chain mRNA
123	CATGGTACTGTGGCT	H728810	23	10	16	15	50		Examples	X87689		H.sapiens mRNA for putative p64 CLCP protein
124	CATGGTCAAAATTC	H737344	0	0	0	10	1		Examples	L12350		Human thrombospondin 2 (THBS2) mRNA
125	CATGCTCTGGGCTT	H752296	25	35	45	76	29		Examples	D21261		Human mRNA (HA1756) for ORF
										D29543		Human keratinocyte cDNA, clone 686
126	CATGCTCTGTGAGAG	H752521	0	5	7	12	2		Examples	H51290		yp07a05.s1 Homo sapiens cDNA clone 186704 3'
										N20338		yx44g12.s1 Homo sapiens cDNA clone 264646 3'
										AA158271		zo76c09.s1 Stragene pancreas (#937208) Homo sapiens cDNA clone 592840 3'
127	CATGCTCTGTGCAGG	H752531	0	0	0	1	13		No Match			
128	CATGCTCTTGAAGCC	H753162	0	1	2	1	10		No Match			
129	CATGGTGAAGGCAGT	H754323	25	14	42	15	89		Examples	X87373		Class C, H.sapiens RPS3a gene
130	CATGGTGAATGACGG	H754567	0	2	8	1	10		Examples	X08058		GLUTATHIONE S-TRANSFERASE P (HUMAN)
131	CATGGTGGGAGGAC	H760361	0	3	2	11	25		Examples	X51439		Human mRNA for serum amyloid A (SAA) protein
132	CATGGTGGTGGAGAA	H761481	2	9	9	13	26		Examples	U15008		Human SnRNP core protein Sm D2 mRNA
133	CATGGTGGAGGGCAC	H762533	1	1	3	6	34		Examples	U62800		Cystatin M (CST6)
134	CATGGTGGTACAGGA	H765003	14	17	15	39	30		Examples	H46430		yo12h12.s1 Homo sapiens cDNA clone 177767 3'
										AA047563		zF13a06.s1 Soares fetal heart NbHH19W Homo sapiens cDNA clone 376786 3'
										AA130701		zo13102.s1 Stragene colon (#937204) Homo sapiens cDNA clone 586779 3'
135	CATGGTTCACCTGCAG	H774629	0	2	1	13	3		Examples	X59288		H.sapiens gene for intercellular adhesion molecule
										M24283		Human major group rhinovirus receptor (HRV) mRNA
										J03132		Human intercellular adhesion molecule-1 (ICAM-1)
										M55100		Human cell surface glycoprotein P3.58 mRNA
136	CATGGTTGTCTTTGG	H781823	1	1	6	30	24		Examples	K02765		Human complement component C3 mRNA, alpha and beta
137	CATGGTTGTGGTTAA	H782013	178	110	14	340	139		Examples	M17987		Human beta-2-microglobulin gene
138	CATGGTTTAAATCGA	H782391	1	6	12	4	14		Examples	D00760		Human mRNA for proteasome subunit HC3
139	CATGTAAGGCTTAAC	H797169	0	0	6	1	12		Examples	X57025		INSULIN-LIKE GROWTH FACTOR IA PRECURSOR (HUMAN)
140	CATGTAATTTTGAA	H802793	0	2	5	2	10		No Match			

	CATGTAATTTTGGAT	H802793							No Match		H.sapiens mRNA for Sm protein G
11	CATGTACATTTTCAT	H806901	1	4	2	3	14		Examples	X83373	
12	CATGTACCCCGTACA	H808370	0	1	4	0	10		No Match		
13	CATGTACCCCTTCTAT	H808925	0	0	0	17	7		No Match		
14	CATGTAGGAAAGTAA	H827437	1	0	5	5	24		Examples	J02931	Human placental tissue factor (two forms) mRNA
										M16353	Human tissue factor mRNA, complete cds
										M27436	Human tissue factor gene, complete cds
15	CATGTAGTTGTCTA	H831416	49	61	89	130			Examples	X64899	H.sapiens mRNA homologous to mouse P21 mRNA.
										X16064	Human mRNA for translationally controlled tumor protein
										L13806	Homo sapiens (clone 04) translationally controlled tumor protein
16	CATGTATATTTTCTC	H839672	1	0	3	8	16		Examples	M98479	Human transglutaminase mRNA
17	CATGTATTTCTGCC	H851834	0	1	2	16	3		Examples	D12149	Human HepG2 3'-directed Mbol cDNA, clone s247
18	CATGTCACAAGCAAA	H856209	10	28	27	24	48		Examples	X80909	H.sapiens alpha NAC mRNA
19	CATGTCCAAATCGAT	H868569	0	1	0	43	17		Examples	X56134	Human mRNA for vimentin.
										Z19554	H.sapiens vimentin gene
										M14144	Human vimentin gene, complete cds
										M25246	Human vimentin (HuVim3) mRNA, 3' end
20	CATGTCCACTGGCCT	H870310	0	0	1	12	2		Examples	N92906	zb57a08.s1 Homo sapiens cDNA clone 307670 3'
										T17488	NIB978 Normalized infant brain, Bento Soares Homo sapiens cDNA 3' end
										AA349906	EST56900 Infant brain Homo sapiens cDNA 3' end
21	CATGTCCATCTGTTG	H871920	6	6	10	25	5		Examples	X67016	H.sapiens mRNA for amphiglycan
										D13292	Human mRNA for ryadocan core protein
22	CATGTCGCTTTTATC	H899060	2	5	15	1	69		Examples	M77233	Human ribosomal protein S7 mRNA
23	CATGTCCTCTGATGCT	H908858	1	5	2	46	19		Examples	S48568	tissue inhibitor of metalloproteinase 2 (3'-end region)
24	CATGCTCTTGTAAGTCTG	H916232	0	4	3	1	13		Examples	N71680	yz93b03.s1 Homo sapiens cDNA clone 290573 3'
25	CATGCTCTTGTCATA	H916372	14	22	15	20	45		Examples	X03083	Human lactate dehydrogenase-A gene
										X02152	Human mRNA for lactate dehydrogenase-A
										X02153	Human pseudogene for lactate dehydrogenase-A
26	CATGTGAAGTCACTG	H920392	1	1	6	0	16		No Match		
27	CATGTGAAGTTATAC	H920525	0	1	3	6	11		Examples	X07979	CTGTGG, Class A, Human mRNA for fibronectin receptor beta subunit.

158	CATGTGATGCTGGT	H932731	0	8	3	11	12	Examples	AA027860	zk05h07.s1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone 469693 3'
159	CATGTGCCATCTGTA	H938876	1	3	7	3	16	Examples	M25753	G2/MITOTIC-SPECIFIC CYCLIN B1 (HUMAN)
									T60151	yc22c04.s1 Homo sapiens cDNA clone 81414 3'
									R67969	yi29g08.s1 Homo sapiens cDNA clone 140702 3'
										zo9103.s1 Stratagene ovarian cancer (#937219) Homo sapiens cDNA clone 594269 3' similar to SW:NGAL_HUMAN P80188 NEUTROPHIL GELATINASE-ASSOCIATED LIPOCALIN PRECURSOR
160	CATGTGCCCTCARRA	H939841	11	13	3	13	43	Examples	AA169614	zb15d08.s1 Homo sapiens cDNA clone 302127 3' similar to SW:NGAL_HUMAN P80188 NEUTROPHIL GELATINASE-ASSOCIATED LIPOCALIN PRECURSOR
161	CATGTGCCCTCAGAA	H939849	3	4	0	11	19	Examples	N79823	zm90h04.s1 Stratagene ovarian cancer (#937219) Homo sapiens cDNA clone 545239 3' similar to SW:NGAL_HUMAN P80188 NEUTROPHIL GELATINASE-ASSOCIATED LIPOCALIN PRECURSOR
162	CATGTGCCCTCAGGA	H939851	13	31	10	25	83	Examples	AA075896	z181e07.s1 Stratagene colon (#937204) Homo sapiens cDNA clone 511044 3'
162	CATGTGCCCTCAGGC	H920392						No Match		
163	CATGTGCCCTTACTTT	H941856	0	3	1	2	12	Examples	AA100279	
164	CATGTGCGCTGGCCC	H944038	2	5	2	17	3	No Match		zk10a01.s1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone 470088 3'
165	CATGTGCTTCATCTG	H949560	2	6	4	16	16	Examples	AA029262	yv66e10.s1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone 247722 3'
									N54281	zn76c02.s1 Stratagene NT2 neuronal precursor 937230 Homo sapiens cDNA clone 564098 3'
									AA114075	Homo sapiens guanylate kinase (GUK1) mRNA
166	CATGTGGAGTGAGG	H953251	18	15	7	22	48	Examples	L76200	Human mRNA for precursor of apolipoprotein C1
167	CATGTGCCCCCAGGT	H955723	0	3	3	37	4	Examples	X00570	Homo sapiens cathepsin B mRNA
168	CATGTGGGTGAGCCA	H962086	13	15	13	76	27	Examples	L16510	Human cathepsin B proteinase mRNA, complete cds
									M14221	Human enigma gene
169	CATGTGTGAGCCCT	H975446	3	3	3	22	8	Examples	L35240	Homo sapiens ribosomal protein L34 (RPL34) mRNA
170	CATGTGTGCTAAATG	H976644	8	21	26	18	50	Examples	L38941	Human gene for histone H1(0)
171	CATGTGTGTGTTTGT	H978687	6	7	16	25	15	Examples	X03473	zk23g08.s1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone 471422 3'
172	CATGTTTATGGATCTC	H997944	0	1	1	21	1	Examples	AA034505	

[illegible]

S1	CATGTTTCCTTCCTT	H1038296	0	6	3	7	17	Examples	M20471	Human brain-type clathrin light-chain a mRNA
									M20472	Human lymphocyte clathrin light-chain A mRNA
S2	CATGTTTGACCTTT	H1041504	2	0	0	16	1	Examples	X78947	H.sapiens mRNA for connective tissue growth factor
									U14750	Human connective tissue growth factor mRNA
S4	CATGTTTGTTAAAA	H1044225							H06492	y178c08.s1 Homo sapiens cDNA clone 44273 3'
									T35952	EST94173 Homo sapiens cDNA 3' end similar to None
									AA253218	zr53g10.s1 Soares NhlMPu S1 Homo sapiens cDNA clone 667170 3'

Table 5 - Transcripts increased in pancreas and colorectal cancer  
 SAGE tag that were elevated in both in colorectal and pancreatic tumor,  
 and are likely to be specific for tumor in general.

	Tag Sequence	Tag Number	Accession	Description
1	CATG TGGAAATGAC C	-950498	M10629	Human alpha-1 collagen gene, 3' end with polyA sit
2	CATG CACTTCRAGG G	-294155	U42376	Human retinolic acid induced RIG-E precursor (E) mR
			U56145	Human thymic shared antigen-1/stem cell antigen-2
3	CATG ATGTGARAGAG T(A)	-243747	J03040	Human SPARC/osteonectin mRNA, complete cds.
			M25746	Human osteonectin gene exon 10, complete cds.
4	CATG GCCCAGGAC C	-610466	X53416	Human mRNA for actin-binding protein (filamin) (AB
5	CATG ATCTTGTAC T	-229106	X02761	Human mRNA for fibronectin (FN precursor).
			K00799	human fibronectin (fn) 3' coding region and flank,
6	CATG GTGCGCTCAG C	-760291	X58536	Human mRNA for HLA class I locus C heavy chain.
			M26432	Human MHC class I HLA-C.1 gene, complete cds.
7	CATG ACAGGCTACG G	-76231	M95787	Human 22kDa smooth muscle protein (SM22) mRNA, com
			M83106	Human SM22 mRNA, 5' end.
8	CATG GTGTGTTTGT A	-769020	M77349	Human transforming growth factor-beta induced gene
9	CATG GATTTCTCAG C	-589267	X53279	Human mRNA for placental-like alkaline phosphatase
			X55958	H.sapiens mRNA for alkaline phosphatase.
			J04948	Human alkaline phosphatase (ALP-1) mRNA, complete
10	CATG ACCATTCTGC T	-85882	X57351	Human 1-8D gene from interferon-inducible gene fam
			X02490	Human interferon-inducible mRNA (cDNA 1-8).
11	CATG TCCTTCTCCA C	-884181	X15804	Human mRNA for alpha-actinin.
12	CATG CTTCTGTGTA C,T	-515821	D80012	Human mRNA for KIAA0190 protein.
13	CATG ATGTAAAAA T	-241665	M74090	Human TB2 gene mRNA, 3' end.
			J03801	Human lysozyme mRNA, complete cds with an Alu repe
			M19045	Human lysozyme mRNA, complete cds.
14	CATG GGCACAGGAC C	-673954	X17620	Human mRNA for Nm23 protein, involved in developme
			X75598	H.sapiens nm23H1 gene.
15	CATG AATATTGAGA A	-53129	U62962	Human Int-6 mRNA, complete cds.
16	CATG TTTTGTATAA A	-1048113	D16891	Human HepG2 3' region cDNA, clone hmd2cl1.
17	CATG CAGCTGGCCA T	-302741	X53743	H.sapiens mRNA for fibulin-1 C.



18	CATG	GTTCACTTA	G	-774461	X00497	Human mRNA for HLA-DR antigens associated invariant
					M13560	Human Ia-associated invariant gamma-chain gene, ex
19	CATG	AAAAGAACT	T	-2056	Y00345	Human mRNA for polyA binding protein.
20	CATG	AATGCAGGCA	G	-58533	M61831	Human S-adenosylhomocysteine hydrolase (AHCY) mRNA
					M61832	Human S-adenosylhomocysteine hydrolase (AHCY) mRNA
21	CATG	TGAATAAAA	C	-918273	X16934	Human hb23 gene for B23 nucleophosmin.
					M28699	Homo sapiens nucleolar phosphoprotein B23 (NPM1) m
					M23613	Human nucleophosmin mRNA, complete cds.
					M26697	Human nucleolar protein (B23) mRNA, complete cds.
22	CATG	TTATGGGATC	T	-998030	M24194	Human MHC protein homologous to chicken B complex
23	CATG	CAATAAATGT	T	-274492	D23661	Human mRNA for ribosomal protein L37, complete cds
					L11567	Homo sapiens ribosomal protein L37 mRNA, complete
						Human mRNA for collagen binding protein 2.
24	CATG	AGCCTTTGTT	G	-155632	D83174	Human 1-8U gene from interferon-inducible gene fam
25	CATG	ACCTGTATCC	C	-97078	X57352	Human acidic ribosomal phosphoprotein P1 mRNA, com
26	CATG	TTCAATAAAA	A	-1000193	M17886	human transcobalamin I mRNA, complete cds.
					J05068	Human apolipoprotein E mRNA, complete cds.
27	CATG	CGACCCACG	C	-398663	M12529	Human apolipoprotein E (epsilon 2 and 3 alleles) m
					K00396	Human UBA52 adrenal mRNA for ubiquitin-52 amino ac
28	CATG	CAGATCTTTG	T	-298495	X56998	Human UBA52 placental mRNA for ubiquitin-52 amino
					X56999	Human DNA inserts showing sperm-specific hypomethy
29	CATG	CTGGCAGCG	C	-501287	X07491	Human ubiquitin carrier protein (E2-EPF) mRNA, com
					M91670	Human prohibitin (PHB) gene, exons 1-7.
30	CATG	ATTGGCTTAA	A	-256497	L14272	prohibitin [human, mRNA, 1043 nt].
					S85655	Human nicotinic acetylcholine receptor alpha6 subu
31	CATG	GTGGTGACA	C	-765573	U62435	Human breast and ovarian cancer susceptibility pro
					U68041	Human parathymosin mRNA, complete cds.
32	CATG	TCCTGCCCA	T	-883029	M24398	H. sapiens RNA for nm23-H2 gene.
33	CATG	ACTGGGCTA	T	-125661	X58965	Human putative NDP kinase (nm23-H2S) mRNA, complet
					M36981	Homo sapiens c-myc transcription factor (puf) mRNA
					L16785	Human ribosomal protein L23a mRNA, partial cds.
34	CATG	AAGAAGATAG	A	-33331	U02032	Human ribosomal protein L23a mRNA, complete cds.
					U37230	Human ribosomal protein L23a mRNA, complete cds.
					U43701	Human ribosomal protein L23a mRNA, complete cds.

				L13799	Homo sapiens (clone 01) liver expressed protein mRNA
35	CATG ACATCATCGA	T	-79065	L06505	Human ribosomal protein L12 mRNA, complete cds.
36	CATG CTGTTGGTGA	T	-507577	D14530	Human homolog of yeast ribosomal protein S28, comp
37	CATG ATTATTTTC	T	-249854	X57959	H. sapiens mRNA for ribosomal protein L7.
				X57958	H. sapiens mRNA for ribosomal protein L7.
				X52967	Human mRNA for ribosomal protein L7.
				L16558	Human ribosomal protein L7 (RPL7) mRNA, complete c
38	CATG GCTTTTAAGG	A	-655115	L06498	Homo sapiens ribosomal protein S20 (RPS20) mRNA, c
39	CATG GGCAAGAAGA	A	-672265	L19527	Homo sapiens ribosomal protein L27 (RPL27) mRNA, c
				L25346	Homo sapiens ribosomal protein L27 (homologue of r
40	CATG CTCCTTCGAGA	A	-490889	Y00433	Human mRNA for glutathione peroxidase (EC 1.11.1.9
				Y00483	Human gene for glutathione peroxidase.
				X13710	H. sapiens unspliced mRNA for glutathione peroxidase
				X13709	Human gpx1 mRNA for glutathione peroxidase.
				M21304	Human glutathione peroxidase (GPX1) mRNA, complete
41	CATG CTGTTGATTG	C	-507455	X04347	Human liver mRNA fragment DNA binding protein UPI
				U00947	Human clone C4E 3.2 (CAC)n/(GTG)n repeat-contains
42	CATG CTGGGTTAAT	A	-502724	M81757	H. sapiens S19 ribosomal protein mRNA, complete cds
43	CATG ATGCTGGTA	T	-239533	X17206	Human mRNA for LLRep3.
44	CATG GATGCTGCCA	A	-583573	X59357	Human mRNA for Epstein-Barr virus small RNAs (EBER
				L21756	Homo sapiens acute myeloid leukemia associated pro
				D17652	Human mRNA for HBp15/L22, complete cds.
				S76343	AML1...EAP (translocation breakpoint) [human, chro
45	CATG CCTTCGAGAT	C	-390692	U14970	Human ribosomal protein S5 mRNA, complete cds.
46	CATG CTCCTCACCT	G	-482584	U16811	Human Bak mRNA, complete cds.
47	CATG TGTGTTGAGA	G	-978825	X16869	Human Bak protein mRNA, complete cds.
				X16872	Human mRNA for elongation factor 1-alpha (clone CE
				X03558	Human DNA for elongation factor 1-alpha (clone lam
				D17182	Human mRNA for elongation factor 1 alpha subunit (
				D17245	Human HepG2 3' region MboI cDNA, clone hmd2h03m3.
				D17259	Human HepG2 3' region MboI cDNA, clone hmd4h05m3.
				D17276	Human HepG2 3' region MboI cDNA, clone hmd5d07m3.
					Human HepG2 3' region MboI cDNA, clone hmd6a12m3.

			M27364	Human elongation factor 1 alpha mRNA, 3' end.
			M29548	Human elongation factor 1-alpha (EF1A) mRNA, parti
			L41490	Homo sapiens oncogene PTF-1 mRNA, complete cds.
			L41498	Homo sapiens oncogene PTF-1 mRNA, complete cds.
				Homo sapiens oncogene PTF-1 mRNA, complete cds.
				Human ribosomal protein L39 mRNA, complete cds.
48	CATG TTACCATATC	A	-988366 U57846	H. sapiens GPx-4 mRNA for phospholipid hydroperoxid
49	CATG GCCTGCTGGG	C	-621035 X71973	H. sapiens gene for ribosomal protein L38.
50	CATG CCTCGGAAAA	T	-383489 Z26876	H. sapiens mRNA for ribosomal protein L6.
51	CATG TACAAGAGGA	A	-803369 X69391	Human mRNA for DNA-binding protein, TAXREB107, com
			-803369 D17554	neoplasm-related C140 product [human, thyroid carc
			-803369 S71022	Human beta-tubulin pseudogene.
52	CATG AACGACCTCG	T	-24951 V00598	Human mRNA fragment encoding beta-tubulin. (from c
			-24951 V00599	Human mRNA for neurite outgrowth-promoting protein
53	CATG CCTGCTCTG	T	-358783 X55110	Human stimulator of TAR RNA binding (SRE) mRNA, co
54	CATG CCCAGGGAGA	A	-346761 U38846	Human HepG2 3' region cDNA, clone hmd4f11.
			D16933	H. sapiens mRNA for elongation factor 2.
55	CATG AGCACCTCCA	G	-148949 Z11692	H. sapiens HRPL4 mRNA.
56	CATG CGCGGAACA	C	-416261 X73974	Human mRNA for ribosomal protein, complete cds.
			D23660	Human 26-kDa cell surface protein TAPA-1 mRNA, com
57	CATG CTAATAAAAA	A	-458753 M33680	Human glycyl-tRNA synthetase mRNA, complete cds.
58	CATG GCCTGATGTG	G	-686319 U09510	Human glycyl-tRNA synthetase mRNA, complete cds.
			U09507	Human T-cell mRNA for glycyl tRNA synthetase, comp
			D30658	Human mRNA for HL23 ribosomal protein homologue.
59	CATG ATTCTCCAGT	A	-253260 X55954	Human mRNA for ribosomal protein L17.
			X52839	H. sapiens mRNA for laminin-binding protein.
60	CATG GAAAAATGGT	T	-524524 X61156	Human mRNA for potential laminin-binding protein (
			X15005	Human 37 kD laminin receptor precursor/p40 ribosom
			U43901	Human colin carcinoma laminin-binding protein mRNA
			J03799	Human laminin receptor (2H5 epitope) mRNA, 5' end.
			M14199	Human mRNA for ribosomal protein L14, complete cds
61	CATG CAGCTCACTG	A	-302367 D87735	Human (clone CTG-B33) mRNA sequence.
			L10376	CAG-isl 7 (trinucleotide repeat-containing sequenc
			S80520	Human ribosomal protein S29 mRNA, complete cds.
62	CATG ATAATTCTTT	G	-200576 U14973	

			L31610	Homo sapiens (clone cor1-lcl5) S29 ribosomal prote
63	CATG AATCCTGTGG A	-55227	Z28407	H.sapiens mRNA for ribosomal protein L8.
64	CATG AATAGGTCCA A	-51925	M64716	Human ribosomal protein S25 mRNA, complete cds.
65	CATG AAAAAAAAAA A (C, G,T)	-1	X83412	H.sapiens B1 mRNA for mucin.
			Z32564	H.sapiens FRGAMMA mRNA (819bp) for folate receptor
			Z32633	H.sapiens FRGAMMA' mRNA for folate receptor (817bp
			X76180	H.sapiens mRNA for lung amiloride sensitive Na+ ch
			U08470	Human FR-gamma' mRNA, complete cds.
			U08471	Human folate receptor 3 mRNA, complete cds.
			U48697	Human mariner-like element-containing mRNA, clone
			D28532	Human mRNA for renal Na+-dependent phosphate cotra
			M55914	Human c-myc binding protein (MBP-1) mRNA, complete
			L06175	Homo Sapiens P5-1 mRNA, complete cds.
			S73775	calmitine-mitochondrial calcium-binding protein (h
			S77393	transcript ch138 (human, RF1, RF48 stomach cancer c
			X60036	H.sapiens mRNA for mitochondrial phosphate carrier
			X79238	H.sapiens mRNA for ribosomal protein L30.
66	CATG CCAGAACAGA C	-335945	L16991	Human thymidylate kinase (CDC8) mRNA, complete cds
			X80822	H.sapiens mRNA for ORF.
67	CATG AAGTGGAGG A	-44683	X52856	Human cyclophilin-related processed pseudogene.
68	CATG CCTAGCTGGA T	-379369	X52857	Human cyclophilin-related processed pseudogene.
			X52854	Human cyclophilin-related processed pseudogene.
			X52851	Human cyclophilin gene for cyclophilin (EC 5.2.1.8
			Y00052	Human mRNA for T-cell cyclophilin.
69	CATG GAACACATCC A	-528694	X63527	H.sapiens mRNA for ribosomal protein L19.
			S56985	ribosomal protein L19 (human, breast cancer cell 1
70	CATG AAGGAGATGG G	-41531	X69181	H.sapiens mRNA for ribosomal protein L31.
			X15940	Human mRNA for ribosomal protein L31.
				H.sapiens SMCX mRNA.
71	CATG AGGCTACGGA A	-171113	Z29650	Human HepG2 3' region MboI cDNA, clone hmd4cl2m3.
			D17233	Human GST pi gene for glutathione S-transferase pi
72	CATG AGGTCCTAGC C	-177610	X08096	

			X06547	Human mRNA for class Pi glutathione S-transferase
			X15480	Human mRNA for anionic glutathione-S-transferase (
			X08058	Human glutathione S-transferase pi gene.
			U12472	Human glutathione S-transferase (GST phi) gene, co
			U21689	Human glutathione S-transferase-Plc gene, complete
			U62589	Human glutathione S-transferase Plc (GSTpic) mRNA,
			M69113	Human fatty acid ethyl ester synthase-III mRNA seq
			M24485	Homo sapiens (clone pHGST-pi) glutathione S-transf
			X69150	H. sapiens mRNA for ribosomal protein S18.
			M96153	Homo sapiens apolipoprotein B gene sequence.
			L06432	Homo sapiens 18S ribosomal protein (HKE3) mRNA seq
			M17885	Human acidic ribosomal phosphoprotein P0 mRNA, com
			L25899	Human ribosomal protein L10 mRNA, complete cds.
			X58125	Human (D9S55) DNA segment containing (TG)24 repeat
			D17268	Human HepG2 3' region MboI cDNA, clone hmd5h09m3.
			M73791	Human novel gene mRNA, complete cds.
			M64241	Human Wilm's tumor-related protein (QM) mRNA, comp
			S35960	laminin receptor homolog [3' region] (human, mRNA
			M17887	Human acidic ribosomal phosphoprotein P2 mRNA, com
			M11147	Human ferritin L chain mRNA, complete cds.
			M12938	Human ferritin light subunit mRNA, partial cds.
			M10119	Human ferritin light subunit mRNA, complete cds.
			X04409	Human mRNA for coupling protein G(s) alpha-subunit
			X04408	Human mRNA for coupling protein G(s) alpha subunit
			X56009	Human GSA mRNA for alpha subunit of GsGTP binding
			X07036	Human mRNA stimulatory GTP-binding protein alpha s
			M21142	Human guanine nucleotide-binding protein alpha-sub
			M14631	Human guanine nucleotide-binding protein G-s, alph
			Z36832	H. sapiens (xs31) mRNA, 835bp.
			K00558	human alpha-tubulin mRNA, complete cds.
			X56494	H. sapiens M gene for M1-type and M2-type pyruvate
			M23725	Human M2-type pyruvate kinase mRNA, complete cds.
			M26252	Human TCB gene encoding cytosolic thyroid hormone-
73	CATG TGGTGTGAG	G	-965603	
74	CATG CTCACATCT	C	-475448	
75	CATG GTGTTAACCA	G	-769045	
76	CATG AGGGCTTCCA	A	-174037	
77	CATG GGATTGGGCC	T	-671654	
78	CATG ATTAACAAG	C	-246019	
79	CATG TGTACCTGTA	A	-968173	
80	CATG TGGCCCCCACC	C	-955718	

81	CATG TAATRAAGGT	G	-798764	X67247	H.sapiens rps8 gene for ribosomal protein S8.
82	CATG GCATAATAGG	T	-602315	X89401	H.sapiens mRNA for large subunit of ribosomal prot
			U14967		Human ribosomal protein L21 mRNA, complete cds.
			U25789		Human ribosomal protein L21 mRNA, complete cds.
			L38826		Homo sapiens L21 ribosomal protein gene, partial c
			X53778		H.sapiens hng mRNA for uracil DNA glycosylase.
83	CATG TACCATCAAT	A	-807748	U34995	Human normal keratinocyte substructure library mRN
			J02642		Human glyceraldehyde 3-phosphate dehydrogenase mRN
			M36164		Human glyceraldehyde-3-phosphate dehydrogenase mRN
			M33197		Human glyceraldehyde-3-phosphate dehydrogenase (GA
			X14957		Human hmgI mRNA for high mobility group protein I.
94	CATG ATTTGTCCCA	G	-260949	X14958	Human hmgI mRNA for high mobility group protein Y.
			M23614		Human HMG-I protein isoform mRNA (HMG-I gene), clon
			M23619		Human HMG-I protein isoform mRNA (HMG-I gene), clon
			L17131		Human high mobility group protein (HMG-I(Y)) gene
			M23615		Human HMG-Y protein isoform mRNA (HMG-I gene), clon
			M23616		Human HMG-Y protein isoform mRNA (HMG-I gene), clon
			M23617		Human HMG-Y protein isoform mRNA (HMG-I gene), clon
			M23618		Human HMG-Y protein isoform mRNA (HMG-I gene), clon
					Human ribosomal protein L27a mRNA, complete cds.
95	CATG GAGGGAGTTT	C	-567488	U14968	Human ribosomal protein L35 mRNA, complete cds.
86	CATG CGCCGCCGGC	T	-416106	U12465	Human ribosomal protein L35 mRNA, complete cds.
87	CATG GTGAACCCCA	ALL	-753749	Z63072	H.sapiens CpG island DNA genomic MseI fragment, cl
88	CATG GTGAACCCCA	ALL	-753749	X16294	Human repetitive DNA containing interspersed repea
89	CATG AAGACAGTGG	C	-33979	X66699	H.sapiens mRNA for ribosomal protein L37a.
			L06499		Homo sapiens ribosomal protein L37a (RPL37A) mRNA,
			L22154		Human ribosomal protein L37a mRNA sequence.
					Human Hums3 mRNA for 40S ribosomal protein s3.
90	CATG CCCCAGCCAG	T	-348755	X55715	Human XP1PO ribosomal protein S3 (rps3) mRNA, comp
			U14990		Human XP2NE ribosomal protein S3 (rps3) mRNA, comp
			U14991		Human XP2NE ribosomal protein S3 (rps3) mRNA, comp
			U14992		Human IMR-90 ribosomal protein S3 (rps3) mRNA, com
			S42658		S3 ribosomal protein (human, colon, mRNA, 826 nt).
					H.sapiens mRNA for protein homologous to elongatio
91	CATG TGGGAAAGC	C	-959498	X63526	H.sapiens mRNA for elongation factor-1-gamma.
				Z11531	H.sapiens mRNA for elongation factor-1-gamma.

			M55409	Human pancreatic tumor-related protein mRNA, 3' en
92	CATG TGAGGGAATA	A	-928269	Human triosephosphate isomerase mRNA, complete cds
93	CATG GAGGACACGA	G	-549145	Human ribosomal protein S28 mRNA, complete cds.
			M58458	Human ribosomal protein S4 (RPS4X) isoform mRNA, c
			M22146	Human scar protein mRNA, complete cds.
94	CATG AACCGGCCA	A	-26261	Homo sapiens macrophage migration inhibitory facto
			L10612	Human glycosylation-inhibiting factor mRNA, comple
			M95775	Homo sapiens macrophage migration inhibitory facto
			L19686	Homo sapiens macrophage migration inhibitory facto
			M25639	Human migration inhibitory factor (MIF) mRNA, comp
95	CATG TGCACGTTTT	C	-935680	Human mRNA for ribosomal protein L32.
			K03002	Human mRNA from chromosome 15 gene with homology t
96	CATG CACAAACGGT	A	-278636	Human ribosomal protein S27 mRNA, complete cds.
			L19739	Homo sapiens metallopeptidase (MPS1) mRNA, compl
97	CATG GGAGTGGACA	T	-667269	Homo sapiens ribosomal protein L18 (RPL18) mRNA, c
98	CATG GCCGAGGAAG	G	-615043	H.sapiens CpG island DNA genomic MseI fragment, cl
			Z57572	H.sapiens CpG island DNA genomic MseI fragment, cl
			Z56073	H.sapiens CpG island DNA genomic MseI fragment, cl
			X53505	Human mRNA for ribosomal protein S12.
99	CATG GGGGAATCG	C	-696375	Human thymosin beta 10 mRNA, complete cds.
			M20259	Human thymosin beta-10 mRNA, complete cds.
100	CATG GCAGCCATCC	G	-599350	Human ribosomal protein L28 mRNA, complete cds.
			D17257	Human HepG2 3' region MboI cDNA, clone hmd5d04m3.
101	CATG TAAGGAGCTG	A	-796831	H.sapiens RPS26 mRNA.
			X69654	H.sapiens mRNA for ribosomal protein S26.
102	CATG GGCAAGCCCC	A	-672342	Human Csa-19 mRNA, complete cds.
			X79239	H.sapiens mRNA for ribosomal protein S13.
			L01124	Human ribosomal protein S13 (RPS13) mRNA, complete
103	CATG GTTCCCTGCG	C	-775658	H.sapiens fau mRNA.
			U02523	Human FAULP pseudogene, trinucleotide repeat regio
104	CATG CCGTCCAAGG	G	-374027	Human ribosomal protein S16 mRNA, complete cds.
	CATG TTGCTCCTCT	G	-1027448	H.sapiens mRNA for homologue to yeast ribosomal pr
			S64030	L41 ribosomal protein homolog [clone 7B6] [human,

105	CATG CAAACCATCC	A	-263478	X12883	Human mRNA for cytokeratin 18.
			X12876		Human mRNA fragment for cytokeratin 18.
			X12881		Human mRNA for cytokeratin 18.
			M24842		Human keratin 18 (K18) gene, complete cds.
			M26325		Human cytokeratin 18 mRNA, 3' end.
			M26326		Human keratin 18 mRNA, complete cds.
			M26327		Human cytokeratin 18 mRNA, 3' end.
106	CATG AGCTCTCCCT	G	-161624	X53777	Human L23 mRNA for putative ribosomal protein.
107	CATG AGTCAGGAG	A(T)	-177315	D86979	Human male bone marrow myeloblast mRNA for KIAA022
			X55923		Human DNA for Alu element PlN6.
			X79699		H. sapiens ALU repeat, 230bp.
			X12544		Human mRNA for HLA class II DR-beta (HLA-DR B).
			Z77989		H. sapiens flow-sorted chromosome 6 HindIII fragmen
			U11831		Human clone 2102V-I chromosome 18p telomeric seque
			U12580		Human Alu repeat sequence A3.
			U12582		Human Alu repeat sequence B2.
			U12583		Human Alu repeat sequence D1.
			U14694		Human Alu-Sb2 repeat, clone HALUSB08.
			U14695		Human Alu-Sb2 repeat, clone HALUSB15.
			U14696		Human Alu-Sb2 repeat, clone HALUSB27.
			U14697		Human Alu-Sb2 repeat, clone HUM-11.
			U14698		Human Alu-Sb2 repeat, clone HSB-8P.
			U14699		Human Alu-Sb2 repeat, clone HUM-9.
			U14700		Human Alu-Sb2 repeat, clone HALUSB35.
			U14701		Human Alu-Sb2 repeat, clone HSB-2P.
			U14704		Human Alu-Sb2 repeat, clone HUM-3.
			U14706		Human Alu-Sb2 repeat, clone HUM-10.
			U14707		Human Alu-Sb2 repeat, clone HUM-7.
			J00120		Human (Lawn) c-myc proto-oncogene, complete coding
			L34653		Homo sapiens platelet/endothelial cell adhesion mo
			M37521		Human XV2c gene.
			S61789		NF1-neurofibromatosis type 1 (deletion breakpoint,
			S73483		phosphorylase kinase catalytic subunit PHKG2 homol



			S75201	cholinesterase (Alu element) [human, Insertion Mut
			S75337	{Y Alu polymorphism, YAP, polymorphic subfamily-3}
108	CATG GGCTGGGT C	-695980	249148	H.sapiens mRNA for ribosomal protein L29.
			U10248	Human ribosomal protein L29 (humpl29) mRNA, compl
			U49083	Human cell surface heparin binding protein HIP mRNA
			D16992	Human HepG2 partial cDNA, clone hmd2d02m5.
			D16911	Human HepG2 3' region cDNA, clone hmd3b09.
			J03537	Human ribosomal protein S6 mRNA, complete cds.
			M20020	Human ribosomal protein S6 mRNA, complete cds.
109	CATG ACGTCTCTT C	-114144		EST
110	CATG TCCTCATCC C	-906438		EST
111	CATG GACTGCGTC C	-555450		EST
112	CATG CTTATCCTG A	-508767		EST
113	CATG GGTGGCAGG G	-719435		EST
114	CATG GCCCTCTGCC A	-613862		EST
115	CATG AACAGAAGCA A	-18469		EST
116	CATG CTGCCGAGCT C	-497192		EST
117	CATG TTCCTCGGGC A	-1007018		EST
118	CATG AACTATACT A	-28872		EST
119	CATG TAGATATGG C	-822331		EST
120	CATG GCCACACCCC A,C	-607318		EST
121	CATG GAACCCCTGGG A	-529899		EST
122	CATG AACTAAAAA A	-28673		EST
123	CATG GAATGTAAG A	-528067		EST
124	CATG ACTCCAAAA A	-119809		EST
125	CATG GTTCGTGCA A	-777109		EST
126	CATG TTACCTCCTT C	-989024		EST
127	CATG GCACAAGAG A	-594051		EST
128	CATG CCCTGGGTC T	-359102		EST
129	CATG GCCTGTATGA G	-621369		EST
130	CATG CCGTCCGGA A	-355689		EST
131	CATG AGGAAGCTG C	-163999		EST
132	CATG TCAGATCTT G	-861056		EST

EST  
EST  
EST  
EST

133	CATG	CCAGGAGGAA	T	-338081
134	CATG	TCACCCACAC	C	-857362
135	CATG	GTGTTGCACA	A	-769605
136	CATG	GCCGTGTCCG	C	-618199

Isolation of partial cDNA (3' fragment) by 3' directed PCR reaction

This procedure is a modification of the protocol described in Polyak et al. (1997) Nature 389:300. Briefly, the procedure uses SAGE tags in PCR reaction such that the resultant PCR product contains the SAGE tag of interest as well as additional cDNA, the length of which is defined by the position of the tag with respect to the 3' end of the cDNA. The cDNA product derived from such a transcript driven PCR reaction can be used for many applications.

RNA from a source believed to express the cDNA corresponding to a given tag is first converted to double-stranded cDNA using any standard cDNA protocol. Similar conditions used to generate cDNA for SAGE library construction can be employed except that a modified oligo-dT primer is used to drive the first strand synthesis. For example, the oligonucleotide of composition 5'-B-TCC GGC GCG CCG TTT T CC CAG TCA CGA(30)-3', contains a poly-T stretch at the 3' end for hybridization and priming from poly-A tails, an M13 priming site for use in subsequent PCR steps, a 5' Biotin label (B) for capture to streptavidin-coated magnetic beads, and an *AscI* restriction endonuclease site for releasing the cDNA from the streptavidin-coated magnetic beads. Theoretically, any sufficiently-sized DNA region capable of hybridizing to a PCR primer can be used as well as any other 8 base pair recognizing endonuclease.

cDNA constructed utilizing this or similar modified oligo-dT primer is then processed exactly as described in U.S. Patent No. (insert) up until adapter ligation where only one adapter is ligated to the cDNA pool. After adapter ligation, the cDNA is released from the streptavidin-coated magnetic beads and is then used as a template for cDNA amplification.

Various PCR protocols can be employed using PCR priming sites within the 3' modified oligo-dT primer and the SAGE tag. The SAGE tag-derived PCR primer employed can be of varying length dictated by 5' extension of the tag into the adaptor sequence. cDNA products are now available for a variety of applications.

This technique can be further modified by: (1) altering the length and/or content of the modified oligo-dT primer; (2) ligating adaptors other than that previously employed within the SAGE protocol; (3) performing PCR from template retained on the streptavidin-coated magnetic beads; and (4) priming first strand cDNA synthesis with non-oligo-dT based primers.

#### Isolation of cDNA using GeneTrapper or modified GeneTrapper Technology

The reagents and manufacturer's instructions for this technology are commercially available from Life Technologies, Inc., Gaithersburg, Maryland. Briefly, a complex population of single-stranded phagemid DNA containing directional cDNA inserts is enriched for the target sequence by hybridization in solution to a biotinylated oligonucleotide probe complementary to the target sequence. The hybrids are captured on streptavidin-coated paramagnetic beads. A magnet retrieves the paramagnetic beads from the solution, leaving nonhybridized single-stranded DNAs behind. Subsequently, the captured single-stranded DNA target is released from the biotinylated oligonucleotide. After release, the cDNA clone is further enriched by using a nonbiotinylated target oligonucleotide to specifically prime conversion of the single-stranded target to double-stranded DNA. Following transformation and plating, typically 20% to 100% of the colonies represent the cDNA clone of interest. To identify the desired cDNA clone, the colonies may be screened by colony hybridization using the <sup>32</sup>P-labeled oligonucleotide as described above for solution hybridization, or alternatively by DNA sequencing and alignment of all sequences obtained from numerous clones to determine a consensus sequence.

The genes which are identified herein as being differentially expressed in normal and cancer cells can be used diagnostically and prognostically. Transcription levels in a test sample suspected of being neoplastic can be determined and compared to the levels in normal colon cells. The test sample may be from any tissue suspected of neoplasia, and particularly from either suspected colorectal or suspected pancreatic cancer cells. The control cells for

the purposes of comparison are normal cells, preferably of the same tissue type as the test sample, *e.g.*, colon cells, or pancreatic duct epithelial cells. Upregulation of transcription or downregulation of transcription is therefore diagnostic of the neoplastic state, depending on what gene is used as a test reagent. Similarly, transcription levels can be monitored to assess patient responses to anti-tumor therapies. Transcription levels will also provide prognostic information. For example, the level of transcription in a test sample can be compared to levels found in *bona fide* normal and tumor cells. More extreme deviations from normal expression levels indicate a poorer prognosis.

Transcription levels can be determined according to any means known in the art. These include, without limitation, Northern blots, nuclear run-on assays, *in vitro* transcription assays, primer extension assays, quantitative reverse transcriptase-polymerase chain reactions (RT-PCR), and hybrid filter binding assays. These techniques are well known in the art. See J.C. Alwine, D.J. Kemp, G.R. Stark, *Proc. Natl. Acad. Sci. U.S.A.* 74, 5350 (1977); K. Zinn, D. Di-Maio, T. Maniatis, *Cell* 34, 865 (1983); G. Veres, R.A. Gibbs, S.E. Scherer, C.T. Caskey, *Science* 237, 415 (1987).

Similarly, upregulated genes and downregulated genes can be detected by measuring expression of their protein products. This can be done by any means known in the art, including but not limited to Western (immuno) blot, enzyme linked immunoadsorbent assay, radioimmunoassay, and enzyme assay. Such techniques are well known in the art. Protein products can be detected in tissue samples of a test patient, using a suspect sample as a test sample, and a matched normal tissue sample from the same tissue type as a control. If normal tissue is not available then a closely related tissue type can be used. Desirably both the samples being compared will be from the same individual. Alternatively, aberrant expression levels of protein products can be detected in body samples, such as blood, serum, feces, urine, sputum. As a control, a normal matched sample can be used from a healthy individual. Aberrant expression levels of transcripts can also be detected in such body samples, particularly in blood and serum.

Probes for use in the assays for transcription levels of particular genes or sets of genes may be RNA or DNA. The probes will be isolated substantially free of other cellular RNAs or DNAs. If the reagent contains one probe then it will comprise at least 50% of the nucleic acids in the reagent composition. If the reagent contains more than one probe, then the proportion will decrease accordingly, so that specific probes will still comprise at least 50% of the nucleic acids in the reagent composition.

Probes can be labeled according to any means known in the art. These may include radioactive labels, fluorescent labels, enzymatic labels, and binding partner labels such as biotin. Means for labeling and detecting probes are well known in the art. Probes comprise at least 10, 11, 12, 15, 20, or 30 contiguous nucleotides of a selected gene.

This invention provides proteins or polypeptides expressed from the polynucleotides of this invention, which is intended to include wild-type and recombinantly produced polypeptides and proteins from procaryotic and eucaryotic host cells, as well as muteins, analogs and fragments thereof. In some embodiments, the term also includes antibodies and anti-idiotypic antibodies.

It is understood that functional equivalents or variants of the wild-type polypeptide or protein also are within the scope of this invention, for example, those having conservative amino acid substitutions. Other analogs include fusion proteins comprising a protein or polypeptide.

The proteins and polypeptides of this invention are obtainable by a number of processes well known to those of skill in the art, which include purification, chemical synthesis and recombinant methods. Full length proteins can be purified from a colon or pancreatic cell or tissue lysate by methods such as immunoprecipitation with antibody, and standard techniques such as gel filtration, ion-exchange, reversed-phase, and affinity chromatography using a fusion protein as shown herein. For such methodology, see for example Deutscher et al. (1999) Guide To Protein Purification: Methods In Enzymology (Vol. 182, Academic Press). Accordingly, this invention also

provides the processes for obtaining these proteins and polypeptides as well as the products obtainable and obtained by these processes.

5 The proteins and polypeptides also can be obtained by chemical synthesis using a commercially available automated peptide synthesizer such as those manufactured by Perkin Elmer/Applied Biosystems, Inc., Model 430A or 431A, Foster City. The synthesized protein or polypeptide can be precipitated and further purified, for example by high performance liquid chromatography (HPLC). Accordingly, this invention also provides a process for chemically synthesizing the proteins of this invention by providing the  
10 sequence of the protein and reagents, such as amino acids and enzymes and linking together the amino acids in the proper orientation and linear sequence.

Alternatively, the proteins and polypeptides can be obtained by well-known recombinant methods as described, for example, in Sambrook et al., (1989), supra, using the host cell and vector systems described above.

15 Also provided by this application are the polypeptides and proteins described herein conjugated to a detectable agent for use in the diagnostic methods. For example, detectably labeled proteins and polypeptides can be bound to a column and used for the detection and purification of antibodies. They also are useful as immunogens for the production of antibodies as  
20 described below. The proteins and fragments of this invention are useful in an in vitro assay system to screen for agents or drugs, which modulate cellular processes.

The proteins of this invention also can be combined with various liquid phase carriers, such as sterile or aqueous solutions, pharmaceutically  
25 acceptable carriers, suspensions and emulsions. Examples of non-aqueous solvents include propyl ethylene glycol, polyethylene glycol and vegetable oils. When used to prepare antibodies, the carriers also can include an adjuvant that is useful to non-specifically augment a specific immune response. A skilled artisan can easily determine whether an adjuvant is required and select one.  
30 However, for the purpose of illustration only, suitable adjuvants include, but

are not limited to Freund's Complete and Incomplete, mineral salts and polynucleotides.

This invention also provides a pharmaceutical composition comprising any of a protein, analog, mutein, polypeptide fragment, antibody, antibody  
5 fragment or anti-idiotypic antibody of this invention, alone or in combination with each other or other agents, and an acceptable carrier. These compositions are useful for various diagnostic and therapeutic methods.

Antibodies can be generated using the proteins encoded by the transcripts identified by the tags disclosed herein. Use of all or portions of the  
10 protein as immunogens is routine in the art. Similarly, fusion proteins can be used as immunogens. Antibodies can be affinity purified using the proteins or portions thereof used as immunogens. Similarly, monoclonal antibodies specifically immunoreactive with the protein sequences of the invention can be generated according to techniques which are well known in the art.

Antibodies can be used analytically to quantitate the expression of  
15 particular transcripts identified herein as upregulated or downregulated in cancer. In addition, antibodies can be conjugated or non-covalently linked to cytotoxic agents, such as cytotoxins, radionuclides, chemotherapeutic drugs, etc. Such antibodies can be used therapeutically to specifically target cancer  
20 cells in which the protein antigens are upregulated. These include the proteins encoded by the transcripts identified by the tags shown in Tables 2, 4, and 5. Means of making such linked cytotoxic antibodies and of administering the same are well known in the art.

Also provided by this invention is an antibody capable of specifically  
25 forming a complex with the proteins or polypeptides as described above. The term "antibody" includes polyclonal antibodies and monoclonal antibodies. The antibodies include, but are not limited to mouse, rat, and rabbit or human antibodies.

Laboratory methods for producing polyclonal antibodies and  
30 monoclonal antibodies, as well as deducing their corresponding nucleic acid sequences, are known in the art, see Harlow and Lane (1988) supra and



Sambrook et al. (1989) *supra*. The monoclonal antibodies of this invention can be biologically produced by introducing protein or a fragment thereof into an animal, e.g., a mouse or a rabbit. The antibody producing cells in the animal are isolated and fused with myeloma cells or heteromyeloma cells to produce hybrid cells or hybridomas. Accordingly, the hybridoma cells producing the monoclonal antibodies of this invention also are provided.

Thus, using the protein or fragment thereof, and well known methods, one of skill in the art can produce and screen the hybridoma cells and antibodies of this invention for antibodies having the ability to bind the proteins or polypeptides.

If a monoclonal antibody being tested binds with the protein or polypeptide, then the antibody being tested and the antibodies provided by the hybridomas of this invention are equivalent. It also is possible to determine without undue experimentation, whether an antibody has the same specificity as the monoclonal antibody of this invention by determining whether the antibody being tested prevents a monoclonal antibody of this invention from binding the protein or polypeptide with which the monoclonal antibody is normally reactive. If the antibody being tested competes with the monoclonal antibody of the invention as shown by a decrease in binding by the monoclonal antibody of this invention, then it is likely that the two antibodies bind to the same or a closely related epitope. Alternatively, one can pre-incubate the monoclonal antibody of this invention with a protein with which it is normally reactive, and determine if the monoclonal antibody being tested is inhibited in its ability to bind the antigen. If the monoclonal antibody being tested is inhibited then, in all likelihood, it has the same, or a closely related, epitopic specificity as the monoclonal antibody of this invention.

The term "antibody" also is intended to include antibodies of all isotypes. Particular isotypes of a monoclonal antibody can be prepared either directly by selecting from the initial fusion, or prepared secondarily, from a parental hybridoma secreting a monoclonal antibody of different isotype by using the sib selection technique to isolate class switch variants using the

procedure described in Steplewski et al. (1985) Proc. Natl. Acad. Sci. 82:8653 or Spira et al. (1984) J. Immunol. Methods 74:307.

This invention also provides biological active fragments of the polyclonal and monoclonal antibodies described above. These "antibody fragments" retain some ability to selectively bind with its antigen or immunogen. Such antibody fragments can include, but are not limited to:

- (1) Fab,
- (2) Fab',
- (3) F(ab')<sub>2</sub>,
- (4) Fv, and
- (5) SCA

A specific example of "a biologically active antibody fragment" is a CDR region of the antibody. Methods of making these fragments are known in the art, see for example, Harlow and Lane, (1988) *supra*.

The antibodies of this invention also can be modified to create chimeric antibodies and humanized antibodies (Oi, et al. (1986) *BioTechniques* 4(3):214). Chimeric antibodies are those in which the various domains of the antibodies' heavy and light chains are coded for by DNA from more than one species.

The isolation of other hybridomas secreting monoclonal antibodies with the specificity of the monoclonal antibodies of the invention can also be accomplished by one of ordinary skill in the art by producing anti-idiotypic antibodies (Herlyn, et al. (1986) *Science* 232:100). An anti-idiotypic antibody is an antibody which recognizes unique determinants present on the monoclonal antibody produced by the hybridoma of interest.

Idiotypic identity between monoclonal antibodies of two hybridomas demonstrates that the two monoclonal antibodies are the same with respect to their recognition of the same epitopic determinant. Thus, by using antibodies to the epitopic determinants on a monoclonal antibody it is possible to identify other hybridomas expressing monoclonal antibodies of the same epitopic specificity.

It is also possible to use the anti-idiotypic technology to produce monoclonal antibodies which mimic an epitope. For example, an anti-idiotypic monoclonal antibody made to a first monoclonal antibody will have a binding domain in the hypervariable region which is the mirror image of the epitope bound by the first monoclonal antibody. Thus, in this instance, the anti-idiotypic monoclonal antibody could be used for immunization for production of these antibodies.

As used in this invention, the term "epitope" is meant to include any determinant having specific affinity for the monoclonal antibodies of the invention. Epitopic determinants usually consist of chemically active surface groupings of molecules such as amino acids or sugar side chains and usually have specific three dimensional structural characteristics, as well as specific charge characteristics.

The antibodies of this invention can be linked to a detectable agent or label. There are many different labels and methods of labeling known to those of ordinary skill in the art.

The antibody-label complex is useful to detect the protein or fragments in a sample, using standard immunochemical techniques such as immunohistochemistry as described by Harlow and Lane (1988) *supra*. Competitive and non-competitive immunoassays in either a direct or indirect format are examples of such assays, e.g., enzyme linked immunoassay (ELISA) radioimmunoassay (RIA) and the sandwich (immunometric) assay. Those of skill in the art will know, or can readily discern, other immunoassay formats without undue experimentation.

The coupling of antibodies to low molecular weight haptens can increase the sensitivity of the assay. The haptens can then be specifically detected by means of a second reaction. For example, it is common to use haptens such as biotin, which reacts avidin, or dinitrophenyl, pyridoxal, and fluorescein, which can react with specific anti-hapten antibodies. See Harlow and Lane (1988) *supra*.

The monoclonal antibodies of the invention also can be bound to many different carriers. Thus, this invention also provides compositions containing the antibodies and another substance, active or inert. Examples of well-known carriers include glass, polystyrene, polypropylene, polyethylene, dextran, nylon, amylases, natural and modified celluloses, polyacrylamides, agaroses and magnetite. The nature of the carrier can be either soluble or insoluble for purposes of the invention. Those skilled in the art will know of other suitable carriers for binding monoclonal antibodies, or will be able to ascertain such, using routine experimentation.

Compositions containing the antibodies, fragments thereof or cell lines which produce the antibodies, are encompassed by this invention. When these compositions are to be used pharmaceutically, they are combined with a pharmaceutically acceptable carrier.

The present invention also provides a screen for various agents which modulate the expression of a gene in a pancreatic or colon cell. To practice the method in vitro, suitable cell cultures or tissue cultures are first provided. The cell can be a cultured cell or a genetically modified cell in which a transcript from SEQ ID NOS:1-732, or their complements, is expressed. Alternatively, the cells can be from a tissue biopsy. The cells are cultured under conditions (temperature, growth or culture medium and gas (CO<sub>2</sub>)) and for an appropriate amount of time to attain exponential proliferation without density dependent constraints. It also is desirable to maintain an additional separate cell culture; one which does not receive the agent being tested as a control.

As is apparent to one of skill in the art, suitable cells may be cultured in microtiter plates and several agents may be assayed at the same time by noting genotypic changes, phenotypic changes or cell death.

When the agent is a composition other than a DNA or RNA, the agent may be directly added to the cell culture or added to culture medium for addition. As is apparent to those skilled in the art, an "effective" amount must be added which can be empirically determined. When the agent is a polynucleotide, it may be directly added by use of a gene gun or

electroporation. Alternatively, it may be inserted into the cell using a gene delivery vehicle or vector as described above.

An agent is a potential therapeutic if it alters the expression of gene in the cell. Altered expression can be detected by assaying for altered mRNA expression or protein expression using the probes, primers and antibodies as described herein.

For the purposes of this invention, an "agent" is intended to include, but not be limited to a biological or chemical compound such as a simple or complex organic or inorganic molecule, a peptide, a protein (e.g. antibody) or a polynucleotide (e.g. anti-sense). A vast array of compounds can be synthesized, for example polymers, such as polypeptides and polynucleotides, and synthetic organic compounds based on various core structures, and these are also included in the term "agent". In addition, various natural sources can provide compounds for screening, such as plant or animal extracts, and the like. It should be understood, although not always explicitly stated that the agent is used alone or in combination with another agent, having the same or different biological activity as the agents identified by the inventive screen. The agents and methods also are intended to be combined with other therapies.

The above disclosure generally describes the present invention. A more complete understanding can be obtained by reference to the following specific examples which are provided herein for purposes of illustration only, and are not intended to limit the scope of the invention.

#### EXAMPLE 1

This example demonstrates the characterization of the general transcription of human colorectal epithelium, colorectal cancers, and pancreatic cancers.

We used the recently developed SAGE (serial analysis of gene expression) method to identify and quantify a total of 303,706 transcripts derived from human colorectal (CR) epithelium, CR cancers or pancreatic cancers (Table 1A ) (3). These transcripts represented approximately 48,741

different genes (4) that ranged in average expression from 1 copy per cell to as many as 5,300 copies per cell (5). The number of different transcripts observed in each cell population varied from 14,247 to 20,471. The bulk of the mRNA mass (75%) consisted of transcripts expressed at more than five copies per cell on average (Table 1B). In contrast, the majority (86%) of transcripts were expressed at less than 5 copies per cell, but in aggregate this low abundance class represented only 25% of the mRNA mass. This distribution was consistently observed among the different samples analyzed and was consistent with previous studies of RNA abundance classes based on RNA-DNA reassociation kinetics (Rot curves). Monte Carlo simulations revealed that our analyses had a 92% probability of detecting a transcript expressed at an average of three copies per cell (7).

Table 1 - Summary of SAGE Analysis

## A. Overall Summary

	Normal		Colon		Colon		Pancreatic		Total
	Colon		Tumors		Cell Lines		Tumors	Cell Lines	
Total Tags	62,168		60,878		60,373		61,592	58,695	303,706
Unique Genes <sup>1</sup>	14,721		19,690		17,092		20,471	14,247	48,741
GenBank <sup>2</sup>	8,753 (59)		10,490 (53)		10,193 (60)		11,547 (56)	8,922 (63)	26,339 (54)

<sup>1</sup> Indicates the number of different genes represented by the total tags analyzed (4).

<sup>2</sup> Indicates the number of genes that matched an entry in GenBank. The number in parentheses indicates the corresponding percentage of total unique tags.

Table 1 - Summary of SAGE Analysis

## B. Summarized by Abundance Classes\*

Copies/Cell	Normal		Colon		Colon		Pancreatic		Pancreatic Cell	
	Colon		Tumors		Cell Lines		Tumors		Lines	Total
> 500										
Unique Genes	62 (29)		54 (25)		54 (19)		32 (11)		70 (26)	55 (19)
GenBank	59 (95)		52 (96)		53 (98)		32 (100)		70 (100)	54 (98)
> 50 and ≤ 500										
Unique Genes	645 (28)		470 (21)		618 (27)		657 (29)		585 (27)	595 (26)
GenBank	545 (84)		429 (91)		579 (94)		609 (93)		529 (90)	553 (93)
> 5 and ≤ 50										
Unique Genes	4,569 (27)		5,011 (29)		5,733 (34)		6,146 (36)		4,895 (31)	6,209 (30)
GenBank	2,893 (63)		3,204 (64)		3,682 (64)		4,054 (66)		3,168 (65)	4,241 (68)



$\leq 5$						
Unique Genes	9,445 (16)	14,155 (25)	10,687 (20)	13,636 (24)	8,697 (16)	41,882 (25)
GenBank	5,256 (56)	6,805 (48)	5,879 (55)	6,852 (50)	5,155 (59)	21,491 (51)

\*For unique genes, the first number denotes the number of different genes (4) represented in the indicated abundance class. The number in parentheses indicates the mass fraction (X100) of total transcripts represented by the indicated abundance class. For GenBank entries, the first number indicates the number of different genes that matched an entry in GenBank in the indicated abundance class. The number in parentheses indicates the corresponding percentage of total genes.

Many of the SAGE tags appeared to represent previously undescribed transcripts, as only 54% of the tags matched entries in GenBank (Table 1). Twenty percent of these matching transcripts corresponded to characterized mRNA sequence entries in GenBank, whereas 80% matched uncharacterized EST entries. As expected, the likelihood of a tag being present in the databases was related to abundance; GenBank matches were identified for 98% of the transcripts expressed at more than 500 copies per cell but for only 51% of the transcripts expressed at  $\leq 5$  copies per cell. Because the SAGE data provide a quantitative assay of transcript abundance, unaffected by differences in cloning or PCR efficiency, these data provide an independent and relatively unbiased estimate of the current completeness of publicly available EST databases.

### EXAMPLE 2

This example demonstrates a comparison of the expression pattern of normal colon epithelium and primary colon cancers.

Comparison of expression patterns between normal colon epithelium and primary colon cancers revealed that the majority of transcripts were expressed at similar levels (Fig. 1A). However, the expression profiles also revealed 289 transcripts that were expressed at significantly different levels [ $P < 0.01$ , (8)]. Of these 289, 181 were decreased in colon tumors compared to normal colon (average decrease 10-fold; Fig. 1B; examples in Fig. 2A). Conversely, 108 transcripts were expressed at higher levels in the colon cancers than in normal colon (average increase 13-fold; Fig. 1C; examples in Fig. 2A). Monte Carlo simulations indicated that the analysis would have detected over 95% of those transcripts expressed at a 6-fold or greater level in normal vs. tumor cells or vice versa (9). Because relatively stringent criteria were used for defining differences [ $P < 0.01$ , (8)], the number of differences reported above is likely to be an underestimate.

EXAMPLE 3

This example demonstrates the similarities and differences between cancer cell line transcription and transcription of primary cancer tissues.

To determine how many of the 289 differences were independent of the cellular microenvironment of cancers *in vivo*, SAGE data from CR cancer cell lines was compared to that from primary CR cancer tissues (Fig. 1B, 1C). Perhaps surprisingly, the majority of transcripts (130 of 181) that were expressed at reduced levels in cancer cells *in vivo* were also expressed at significantly lower levels in the cell lines (Fig. 1B). Likewise, a significant fraction of the transcripts expressed at increased levels in primary cancers were also expressed at higher levels in the CR cancer cell lines (Fig. 1C). Thus, many of the gene expression differences that distinguish normal from tumor cells *in vivo* persist during *in vitro* growth. However, despite these similarities there were also many differences. For example, only 47 of 228 genes expressed at higher levels in CR cancer cell lines were also expressed at high levels in the primary CR cancers.

In combination, comparing the expression pattern of CR cancer cells (in *vivo* or *in vitro*) to normal colon revealed 548 differentially expressed transcripts (Fig. 1B,C, Tables 2 and 3). The average difference in expression for these transcripts was 15 fold. Although the ability to detect differences is influenced by the magnitude of the variance with the power to detect smaller differences being less, 92 transcripts that were less than three fold different were identified among the 548 transcripts. However, those genes exhibiting the greatest differences in expression are likely to be the most biologically important.

EXAMPLE 4

This example demonstrates the similarities and differences between colorectal cancer transcription and pancreatic cancer transcription.

To determine whether the changes noted in CR cancers were neoplasia or cell type specific, we performed SAGE on mRNA derived from pancreatic cancers. A total of 404 transcripts were expressed at higher levels in pancreatic cancers compared to normal colon epithelium (examples in Fig. 2B). The majority (268) of these transcripts were pancreas-specific (10) (Example in Fig. 2C) although 136 were also expressed at high levels in CR cancers. These 136 transcripts constituted 47% of the 289 transcripts increased in CR cancers relative to normal colon and are likely to be related to the neoplastic process rather than to the specific cell type of origin.

EXAMPLE 5

This example demonstrates the reproducibility of the transcription patterns observed among a larger number of cancer samples.

One question that arose from these data is the potential heterogeneity of expression between individual tumors. The SAGE data were acquired from two examples of each tissue type (normal colon, primary CR cancer, CR cancer cell line, etc.). To examine the generality of these expression profiles, we arbitrarily selected 27 differentially expressed transcripts and evaluated them in six to twelve samples of normal colon and primary cancers by Northern blot analysis (11). In general, expression patterns were very reproducible among different samples. Of 10 genes with elevated expression in normal colon relative to CR cancers as determined by SAGE, each was detected in the normal colon samples and was expressed at considerably lower levels in tumors (examples in Fig. 2A). Similarly, most of the genes identified by SAGE as increased in CR or pancreatic cancers were confirmed to be reproducibly expressed in the majority of primary cancers examined by Northern blot (examples in Fig. 2A). It is important to note, however, that there were differences among the cancers, with a few cancers exhibiting particularly high or low levels of individual transcripts. Such differences in gene expression

undoubtedly contribute to the observed heterogeneity in biological properties of cancers derived from the same organ .

#### EXAMPLE 6

5 This example demonstrates the identities of some of the transcripts which were found to be differentially expressed in tumor and normal tissues. What are the identities of the differentially expressed genes? Of the 548 differentially expressed transcripts, 337 were tentatively identified through database comparisons. When tested, the great majority (93%) of these identifications proved to be legitimate (13), as expected from previous SAGE  
10 analyses . Although a large number of differentially expressed genes were identified, some simple patterns did emerge. For example, genes that were expressed at higher levels in normal colon epithelium than in CR tumors were often differentiation-related. These genes included liver fatty acid binding protein , cytokeratin 20 , carbonic anhydrase , guanylin and uroguanylin ,  
15 which are known to be important for the normal physiology or architecture of the colon epithelium (Table 2). On the other hand, genes that were increased in CR cancers were often related to the robust growth characteristics that these cells exhibit. For example, gene products associated with protein synthesis, including 48 ribosomal proteins, five elongation factors, and five genes  
20 involved in glycolysis were observed to be elevated in both CR and pancreatic cancers compared to normal colon cells. Although the majority of the transcripts could not have been predicted to be differentially expressed in cancers, several have previously been shown to be dysregulated in neoplastic cells. The latter included IGFII , B23 nucleophosmin, the Pi form of  
25 glutathione S-transferase, and several ribosomal proteins which were all increased in cancer cells as previously reported. Likewise, Dra and gelsolin were both decreased in cancer as previously reported. Surprisingly, two widely studied oncogenes, *c-fos* and *c-erbB3*, were expressed at much higher levels in normal colon epithelium than CR cancers, in contrast to their up-regulation in  
30 transformed cells .

In summary, these data provide basic information necessary for understanding the gene expression differences that underlie cancer phenotypes. They additionally provide a necessary framework for interpreting the significance of individual differentially expressed genes. Although this study demonstrated that a large number of such differences exist (approximately 500 at the depth of analysis employed), it was equally remarkable that the fraction of transcripts exhibiting significant differences was relatively small, representing 1.5 % of the transcripts detected in any given cell type (26). The fact that many, but not all, of the differences were preserved during in vitro culture demonstrates the utility of cultured lines for examination of some aspects of gene expression, but also provides a note of caution in relying on such lines to perfectly mimic tumors in their natural environment. Finally, the finding that hundreds of specific genes are expressed at different levels in CR cancers, and that some of these are also expressed differentially in pancreatic cancers, provides a wealth of new reagents for future biologic and diagnostic experimentation.

REFERENCES AND NOTES

1. M. D. Adams, *et al.*, *Nature* **377**, *supp.* **28**, 3 (1995); M. Schena, D. Shalon, R. W. Davis, P. O. Brown, *Science* **270**, 467 (1995); J. Derisi, *et al.*, *Nature Genetics* **14**, 457 (1996); T. M. Gress, *et al.*, *Oncogene* **13**, 1819 (1996); D. J. Lockhart, *et al.*, *Nature Biotechnology* **14**, 1675 (1996); M. Schena, *et al.*, *Proc Natl Acad Sci USA* **93**, 10614 (1996).

2. V. E. Velculescu, L. Zhang, B. Vogelstein, K. W. Kinzler, *Science* **270**, 484 (1995); V. E. Velculescu, *et al.*, *Cell* **88**, 243 (1997).

3. To minimize individual variation, approximately equal numbers of tags (30,000) were derived from two different patients for each tissue. For primary tumors (two CR carcinomas and two pancreatic adenocarcinomas), RNA was isolated from portions of tumors judged to contain 60%-90% tumor cells by histopathology. The cells grown in vitro were derived from CR (SW837, Caco2) and pancreatic (ASPC-1, PL45) cancer cell lines. CR epithelial cells were isolated from sections of normal colon mucosa from two patients using EDTA as previously described [S. Nakamura, I. Kino, S. Baba, *Gut* **34**, 1240 (1993)]. Histopathology confirmed that the isolated cells were greater than 90% epithelial. Isolation of Poly-A RNA and SAGE was performed as previously described (2). SAGE data was analyzed by means of SAGE software and GenBank Release 95 as previously described (2).

4. A total of 69,393 different SAGE tags were identified among the 303,706 tags analyzed. A small fraction of these different tags were likely due to sequencing errors. SAGE analysis of yeast (2), wherein the entire genomic sequence is known, demonstrated a sequencing error rate of ~ 0.7%, translating to a SAGE tag error rate of 6.8% ( $1 - 0.993^{10}$ ). Because these sequencing mistakes are essentially random, they do not substantially affect the analysis although they could artificially inflate the number of unique genes identified. Therefore, to be conservative, we reduced our estimate of unique genes identified by this maximum tag error rate (e.g., 6.8% of 303,706 total tags). The number of different tags derived from the same gene due to alternative splicing was assumed to be negligible.

5. Abundances can be simply determined by dividing the observed number of tags for a given transcript by the total number of tags obtained. An estimate of approximately 300,000 transcripts per cell was used to convert the abundances to copies per cell [N. D. Hastie, J. O. Bishop, *Cell* 9, 761 (1976)].

5 6. J. O. Bishop, J. G. Morton, M. Rosbash, M. Richardson, *Nature* 250, 199 (1974); B. Lewin, *Gene Expression Vol 2* (John Wiley and sons, New York 1980).

7. Computer simulations indicated that analysis of 300,000 tags would yield a 92 % chance of detecting a tag for a transcript whose expression was at least three copies per cell on average among the tissues examined and assuming 300,000 transcripts per cell.

8. To minimize the number of assumptions and to account for the large number of comparisons being made, Monte Carlo analysis was used for determining statistical significance. The null hypothesis was that the level, kind, and distribution of transcripts were the same for cancer and normal cells. For each transcript, 100,000 simulations were performed to determine the relative likelihood due to chance alone ("p-chance") of obtaining a difference in expression equal to or greater than the observed difference, given the null hypothesis. This likelihood was converted to an absolute probability value by simulating 40 experiments in which a representative number of transcripts (27,993 transcripts in each experiment) was identified and compared. The distribution of transcripts used for these simulations was derived from the average level of expression observed in the original samples. The distribution of the p-chance scores obtained in the 40 simulated experiments (false positives) was then compared to those obtained experimentally. Based on this comparison, a maximum value of 0.0005 was chosen for p-chance. This yielded a false positive rate that was no higher than 0.01 for the least significant p-chance value below the cutoff.

9. Two hundred simulations assuming an abundance of 0.0001 in one sample and 0.0006 in a second sample revealed a significant difference ( $P < 0.01$ , [8]) 95% of the time.



10. It is not possible to obtain pancreatic ductal epithelium, from which pancreatic carcinomas arise, in sufficient quantities to perform SAGE. It is therefore not possible to determine whether these transcripts were derived from genes that were highly expressed only in pancreatic cancers or were also expressed in pancreatic duct cells.

11. Total RNA isolation and Northern blot analysis was performed as described [ W. S. el-Deiry, *et al.*, *Cell* 75, 817 (1993)].

12. A. H. Owens, D. S. Coffey, S. B. Baylin, Eds., *Tumor Cell Heterogeneity: Origins and Implications* (Academic Press, New York, 1982).

13. Northern blot analyses were done on 45 of the 337 differentially expressed transcripts with tentative database matches. In three cases, the pattern of expression was not differentially expressed as predicted by SAGE and, for the purposes of this calculation, were presumed to represent incorrect database matches.

14. D. C. Rubin, D. E. Ong, J. I. Gordon, *Proc Natl Acad Sci U S A* 86, 1278 (1989); K. Okubo, J. Yoshii, H. Yokouchi, M. Kameyama, K. Matsubara, *DNA Res* 1, 37 (1994).

15. R. Moll, *et al.*, *Differentiation* 53, 75 (1993).

16. J. Sowden, S. Leigh, I. Talbot, J. Delhanty, Y. Edwards, *Differentiation* 53, 67 (1993).

17. F. J. de Sauvage, *et al.*, *Proc Natl Acad Sci U S A* 89, 9089 (1992).

18. R. C. Wiegand, *et al.*, *FEBS Lett* 311, 150 (1992).

19. J. V. Tricoli, *et al.*, *Cancer Res* 46, 6169 (1986); S. Lambert, J. Vivario, J. Boniver, R. Gol-Winkler, *Int J Cancer* 46, 405 (1990).

20. W. Y. Chan, *et al.*, *Biochemistry* 28, 1033 (1989).

21. J. D. Hayes, D. J. Pulford, *Crit Rev Biochem Mol Biol* 30, 445 (1995).

22. G. F. Barnard, *et al.*, *Cancer Res* 52, 3067 (1992); P. J. Chiao, D. M. Shin, P. G. Sacks, W. K. Hong, M. A. Tainsky, *Mol Carcinog* 5, 219 (1992); N. Kondoh, C. W. Schweinfest, K. W. Henderson, T. S. Papas,

*Cancer Res* 52, 791 (1992); G. F. Barnard, *et al.*, *Cancer Res* 53, 4048 (1993); M. G. Denis, *et al.*, *Int J Cancer* 55, 275 (1993); J. M. Frigerio, *et al.*, *Hum Mol Genet* 4, 37 (1995).

23. C. W. Schweinfest, K. W. Henderson, S. Suster, N. Kondoh,  
5 T. S. Papas, *Proc Natl Acad Sci USA* 90, 4166 (1993).

24. M. Tanaka, *et al.*, *Cancer Res* 55, 3228 (1995); D. Medina, F.  
S. Kittrell, C. J. Oborn, M. Schwartz, *Cancer Res* 53, 668 (1993).

25. A. D. Miller, T. Curran, I. M. Verma, *Cell* 36, 51 (1984); M.  
10 H. Kraus, W. Issing, T. Miki, N. C. Popescu, S. A. Aaronson, *Proc Natl Acad Sci USA* 86, 9193 (1989).

26. In the case of normal and neoplastic colon cancer tissue, 548  
differentially transcripts were identified among the 36,125 unique transcripts.

27. All references cited are hereby incorporated by reference herein.

28. Sequences tags in Tables 2-4 are consecutively numbered to  
15 form SEQ ID NOS: 1-732.

CLAIMS

1. A method of diagnosing colon cancer in a sample suspected of being neoplastic, comprising the steps of:

5 comparing the level of at least one transcript in a first sample of a tissue to a second sample, wherein the first sample is of a colonic tissue suspected of being neoplastic and the second sample is of a normal human colonic tissue, and wherein the transcript is identified by a tag selected from the group consisting of those shown in Table 3;

10 identifying the first sample as neoplastic when the level of the at least one transcript is found to be lower in the first sample than in the second sample.

2. A method of diagnosing colon cancer in a sample suspected of being neoplastic, comprising the steps of:

15 comparing the level of at least one transcript in a first sample of a tissue to a second sample, wherein the first sample is of a colonic tissue suspected of being neoplastic and the second sample is of a normal human colonic tissue, and wherein the transcript is identified by a tag selected from the group consisting of those shown in Table 2;

20 identifying the first sample as neoplastic when the level of the at least one transcript is found to be higher in the first sample than in the second sample.

3. The method of claim 1 wherein a comparison of at least two of said transcripts is performed.

25 4. The method of claim 2 wherein a comparison of at least two of said transcripts is performed.

5. The method of claim 1 wherein a comparison of at least five of said transcripts is performed.

6. The method of claim 2 wherein a comparison of at least five of said transcripts is performed.

5 7. The method of claim 1 wherein a comparison of at least ten of said transcripts is performed.

8. The method of claim 2 wherein a comparison of at least ten of said transcripts is performed.

10 9. The method of claim 1 wherein a comparison of at least twenty of said transcripts is performed.

10. The method of claim 2 wherein a comparison of at least twenty of said transcripts is performed.

11. The method of claim 1 wherein a comparison of at least thirty of said transcripts is performed.

15 12. The method of claim 2 wherein a comparison of at least thirty of said transcripts is performed.

13. An isolated and purified human nucleic acid molecule which comprises a SAGE tag selected from SEQ ID NO:1-732.

14. The nucleic acid molecule of claim 13 which is a cDNA molecule.

15. The nucleic acid molecule of claim 13 wherein the SAGE tag is located at the 3' end of the molecule, adjacent to the 3'-most NlaIII restriction enzyme site.

5 16. An isolated nucleotide probe comprising at least 10 nucleotides of a human nucleic acid molecule, wherein the human nucleic acid molecule comprises a SAGE tag selected from SEQ ID NO: 1-732.

17. The probe of claim 16 which comprises the selected SAGE tag.

18. A diagnostic reagent for evaluating neoplasia of a colorectal tissue, comprising at least 2 probes according to claim 16.

10 19. The diagnostic reagent of claim 18 which comprises at least 5 probes according to claim 16.

20. The diagnostic reagent of claim 18 which comprises at least 10 probes according to claim 16.

15 21. The diagnostic reagent of claim 18 which comprises at least 20 probes according to claim 16.

22. The diagnostic reagent of claim 18 which comprises at least 30 probes according to claim 16.

23. A diagnostic reagent for evaluating neoplasia of a colorectal tissue, comprising at least 2 probes according to claim 17.

20 24. A method of diagnosing pancreatic cancer in a sample suspected of being neoplastic, comprising the steps of:

comparing the level of at least one transcript in a first sample of a tissue to a second sample, wherein the first sample is of a pancreatic tissue suspected of being neoplastic and the second sample is of a normal human colon tissue, wherein said transcript is identified by a tag selected from the group consisting of those shown Table 4;

identifying the first sample as neoplastic when the level of the at least one transcript is found to be higher in the first sample than in the second sample.

25. A method of diagnosing cancer in a sample suspected of being neoplastic, comprising the steps of:

comparing the level of at least one transcript in a first sample of a tissue to a second sample, wherein the first sample is of a tissue suspected of being neoplastic and the second sample is of a normal human tissue of the same tissue type, wherein said transcript is identified by a tag selected from the group consisting of those shown Table 5;

identifying the first sample as neoplastic when the level of the at least one transcript is found to be higher in the first sample than in the second sample.

26. A method to aid in the determination of a prognosis for a colon cancer patient, comprising the steps of:

comparing the level of at least one transcript in a first sample of a tissue to a second sample, wherein the first sample is of a neoplastic colonic tissue and the second sample is of a normal human colonic tissue, and wherein the transcript is identified by a tag selected from the group consisting of those shown in Table 3;

determining a poorer prognosis if the level of the at least one transcript is found to be lower in the first sample than in the second sample.

27. A method to aid in determining a prognosis for a patient with colon cancer, comprising the steps of:

comparing the level of at least one transcript in a first tissue sample to a second sample, wherein the first sample is of a colonic cancer tissue and the second sample is of a normal human colonic tissue, and wherein the transcript is identified by a tag selected from the group consisting of those shown in Table 2;

determining a poorer prognosis if the level of the at least one transcript is found to be higher in the first sample than in the second sample.

28. A method of diagnosing colon cancer in a sample suspected of being neoplastic, comprising the steps of:

comparing the level of expression of at least one protein in a first sample of a tissue to a second sample, wherein the first sample is of a colonic tissue suspected of being neoplastic and the second sample is of a normal human colonic tissue, and wherein the protein is encoded by a transcript identified by a tag selected from the group consisting of those shown in Table 3;

identifying the first sample as neoplastic when the level of expression of the protein is found to be lower in the first sample than in the second sample.

29. A method of diagnosing colon cancer in a sample suspected of being neoplastic, comprising the steps of:

comparing the level of expression of at least one protein in a first sample of a tissue to a second sample, wherein the first sample is of a colonic tissue suspected of being neoplastic and the second sample is of a normal human colonic tissue, and wherein the protein is encoded by a transcript identified by a tag selected from the group consisting of those shown in Table 2;

identifying the first sample as neoplastic when expression of the protein is found to be higher in the first sample than in the second sample.

30. A method to aid in determining a prognosis of a patient having pancreatic cancer, comprising the steps of:

5 comparing the level of at least one transcript in a first sample of a tissue to a second sample, wherein the first sample is of a neoplastic pancreatic tissue and the second sample is of a normal human colon tissue, wherein said transcript is identified by a tag selected from the group consisting of those shown Table 4;

determining a poorer prognosis if the level of the at least one transcript is found to be higher in the first sample than in the second sample.

10 31. A method to aid in providing a prognosis for a cancer patient, comprising the steps of:

15 comparing the level of at least one transcript in a first sample of a tissue to a second sample, wherein the first sample is of a neoplastic tissue and the second sample is of a normal human tissue of the same tissue type, wherein said transcript is identified by a tag selected from the group consisting of those shown Table 5;

determining a poorer prognosis if the level of the at least one transcript is found to be higher in the first sample than in the second sample.

20 32. A method of diagnosing pancreatic cancer in a sample suspected of being neoplastic, comprising the steps of:

25 comparing the level of expression of at least one protein encoded by a transcript in a first sample of a tissue to a second sample, wherein the first sample is of a pancreatic tissue suspected of being neoplastic and the second sample is of a normal human colon tissue, wherein said protein is encoded by a transcript identified by a tag selected from the group consisting of those shown Table 4;

identifying the first sample as neoplastic when expression of the protein is found to be higher in the first sample than in the second sample.



33. A method of diagnosing cancer in a sample suspected of being neoplastic, comprising the steps of:

5                   comparing the level of expression of at least one protein in a first sample of a tissue to a second sample, wherein the first sample is of a tissue suspected of being neoplastic and the second sample is of a normal human tissue, wherein said protein is encoded by a transcript identified by a tag selected from the group consisting of those shown Table 5;

                  identifying the first sample as neoplastic when expression of the protein is found to be higher in the first sample than in the second sample.

10           34. A method to aid in the determination of a prognosis for a colon cancer patient, comprising the steps of:

                  comparing the level of expression of at least one protein in a first sample of a tissue to a second sample, wherein the first sample is of a neoplastic colonic tissue and the second sample is of a normal human colonic tissue, and wherein the protein is encoded by a transcript identified by a tag selected from the group consisting of those shown in Table 3;

15                   determining a poorer prognosis if the level of expression is found to be lower in the first sample than in the second sample.

20           35. A method to aid in determining a prognosis for a patient with colon cancer, comprising the steps of:

                  comparing the level of expression of at least one protein in a first tissue sample to a second sample, wherein the first sample is of a colonic cancer tissue and the second sample is of a normal human colonic tissue, and wherein the protein is encoded by a transcript identified by a tag selected from the group consisting of those shown in Table 2;

25                   determining a poorer prognosis if the level of expression is found to be higher in the first sample than in the second sample.

36. A method to aid in determining a prognosis of a patient having pancreatic cancer, comprising the steps of:

5           comparing the level of expression of at least one protein in a first sample of a tissue to a second sample, wherein the first sample is of a neoplastic pancreatic tissue and the second sample is of a normal human colon tissue, wherein said protein is encoded by a transcript identified by a tag selected from the group consisting of those shown Table 4;

          determining a poorer prognosis if the level of expression is found to be higher in the first sample than in the second sample.

10       37. A method to aid in providing a prognosis for a cancer patient, comprising the steps of:

          comparing the level of expression of at least one protein in a first sample of a tissue to a second sample, wherein the first sample is of a neoplastic tissue and the second sample is of a normal human tissue of the same  
15       tissue type, wherein said protein is encoded by a transcript identified by a tag selected from the group consisting of those shown Table 5;

          determining a poorer prognosis if the level of expression is found to be higher in the first sample than in the second sample.

38. A method of treating a cancer cell, comprising the step of:

20           administering to a cancer cell an antibody which specifically binds to a protein encoded by a transcript identified by a tag selected from the group consisting of those shown in Tables 2, 4, and 5, wherein the antibody is linked to a cytotoxic agent.

25       39. An antibody linked to a cytotoxic agent, wherein the antibody specifically binds to a protein encoded by a transcript identified by a tag selected from the group consisting of those shown in Tables 2, 4, and 5.

40. A method of detecting colon cancer in a patient, comprising the steps of:

5                   comparing the level of at least one protein in a first body sample to a second body sample, wherein the first sample is a body sample of the patient and the second sample is of a normal human, wherein the protein is encoded by a transcript identified by a tag selected from the group consisting of those shown in Table 2, wherein the first and second body sample is a sample selected from the group consisting of blood, urine, feces, sputum, and serum;

10                   identifying neoplasia when the level of the at least one protein is found to be higher in the first sample than in the second sample.

41. A method of detecting pancreatic cancer in a patient, comprising the steps of:

15                   comparing the level of at least one protein encoded by a transcript in a first sample of a tissue to a second sample, wherein the first sample is of the patient and the second sample is of a normal human, wherein said protein is encoded by a transcript identified by a tag selected from the group consisting of those shown Table 4, wherein the first and second sample is a sample selected from the group consisting of blood, urine, feces, sputum, and serum;

20                   identifying neoplasia when the level of the at least one protein is found to be higher in the first sample than in the second sample.

42. A method of detecting cancer in a patient, comprising the steps of:

25                   comparing the level of at least one protein in a first sample to a second sample, wherein the first sample is of patient and the second sample is of a normal human, wherein said protein is encoded by a transcript identified by a tag selected from the group consisting of those shown Table 5, wherein the first and second body sample is a sample selected from the group consisting of blood, urine, feces, sputum, and serum;

identifying neoplasia when the level of the at least one protein is found to be higher in the first sample than in the second sample.

43. A method to aid in determining a prognosis for a patient with colon cancer, comprising the steps of:

5                   comparing the level of at least one protein in a first sample to a second sample, wherein the first sample is of a colonic cancer patient and the second sample is of a normal human, wherein the protein is encoded by a transcript identified by a tag selected from the group consisting of those shown in Table 2, wherein the first and second sample is a sample selected from the  
10                   group consisting of blood, urine, feces, sputum, and serum;

                  determining a poorer prognosis if the level of the at least one protein is found to be higher in the first sample than in the second sample.

44. A method to aid in determining a prognosis of a patient having pancreatic cancer, comprising the steps of:

15                   comparing the level of at least one protein in a first sample to a second sample, wherein the first sample is of a pancreatic cancer patient and the second sample is of a normal human, wherein said protein is encoded by a transcript identified by a tag selected from the group consisting of those shown Table 4, wherein said first and second sample is a sample selected from the  
20                   group consisting of blood, urine, feces, sputum, and serum;

                  determining a poorer prognosis if the level of the at least one protein is found to be higher in the first sample than in the second sample.

45. A method to aid in providing a prognosis for a cancer patient, comprising the steps of:

25                   comparing the level of expression of at least one protein in a first sample to a second sample, wherein the first sample is of a cancer patient and the second sample is of a normal human, wherein said protein is encoded by a transcript identified by a tag selected from the group consisting of those

shown Table 5, wherein the first and second sample is a sample selected from the group consisting of blood, urine, feces, sputum, and serum;

determining a poorer prognosis if the level of the at least one protein is found to be higher in the first sample than in the second sample.

5           46.     A method of detecting colon cancer in a patient, comprising the steps of:

10                     comparing the level of at least one transcript in a first body sample to a second body sample, wherein the first sample is a body sample of the patient and the second sample is of a normal human, wherein the transcript is identified by a tag selected from the group consisting of those shown in Table 2, wherein the first and second body sample is a sample selected from the group consisting of blood, urine, feces, sputum, and serum;

                   identifying neoplasia when the level of the at least one transcript is found to be higher in the first sample than in the second sample.

15           47.     A method of detecting pancreatic cancer in a patient, comprising the steps of:

20                     comparing the level of at least one transcript in a first sample of a tissue to a second sample, wherein the first sample is of the patient and the second sample is of a normal human, wherein said transcript is identified by a tag selected from the group consisting of those shown Table 4, wherein the first and second sample is a sample selected from the group consisting of blood, urine, feces, sputum, and serum;

                   identifying neoplasia when the level of the at least one transcript is found to be higher in the first sample than in the second sample.

25           48.     A method of detecting cancer in a patient, comprising the steps of:

                   comparing the level of at least one transcript in a first sample to a second sample, wherein the first sample is of patient and the second sample is of a normal human, wherein said transcript is identified by a tag selected

from the group consisting of those shown Table 5, wherein the first and second body sample is a sample selected from the group consisting of blood, urine, feces, sputum, and serum;

identifying neoplasia when the level of the at least one transcript is found to be higher in the first sample than in the second sample.

49. A method to aid in determining a prognosis for a patient with colon cancer, comprising the steps of:

comparing the level of at least one transcript in a first sample to a second sample, wherein the first sample is of a colonic cancer patient and the second sample is of a normal human, wherein the transcript is identified by a tag selected from the group consisting of those shown in Table 2, wherein the first and second sample is a sample selected from the group consisting of blood, urine, feces, sputum, and serum;

determining a poorer prognosis if the level of the at least one transcript is found to be higher in the first sample than in the second sample.

50. A method to aid in determining a prognosis of a patient having pancreatic cancer, comprising the steps of:

comparing the level of at least one transcript in a first sample to a second sample, wherein the first sample is of a pancreatic cancer patient and the second sample is of a normal human, wherein said transcript is identified by a tag selected from the group consisting of those shown Table 4, wherein said first and second sample is a sample selected from the group consisting of blood, urine, feces, sputum, and serum;

determining a poorer prognosis if the level of the at least one transcript is found to be higher in the first sample than in the second sample.

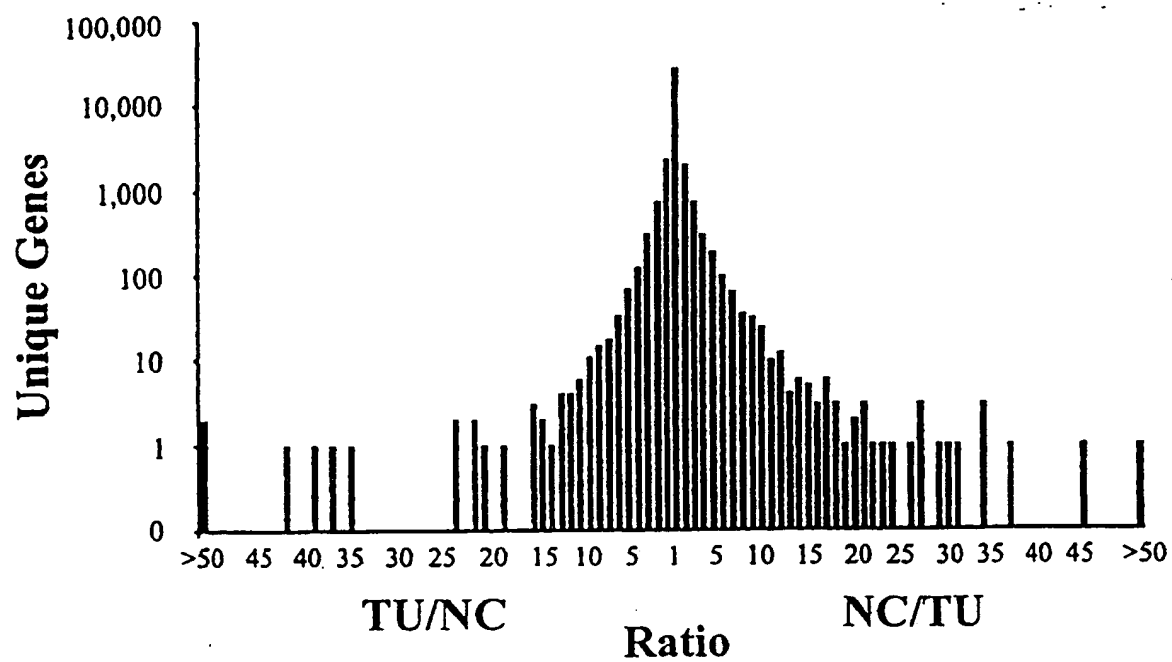
51. A method to aid in providing a prognosis for a cancer patient, comprising the steps of:

5                   comparing the level of expression of at least one transcript in  
a first sample to a second sample, wherein the first sample is of a cancer patient  
and the second sample is of a normal human, wherein said transcript is  
identified by a tag selected from the group consisting of those shown Table 5,  
wherein the first and second sample is a sample selected from the group  
consisting of blood, urine, feces, sputum, and serum;

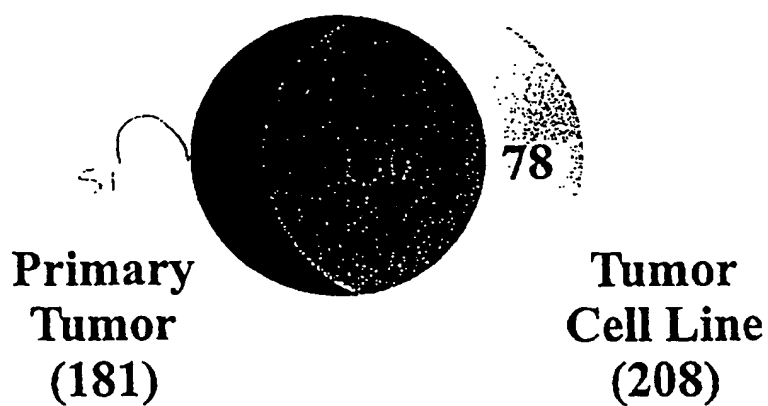
                  determining a poorer prognosis if the level of the at least one  
transcript is found to be higher in the first sample than in the second sample.

10           52.    A method for screening for candidate agents that modulate the  
expression of a polynucleotide selected from the group consisting of the  
polynucleotides in SEQ ID NOS:1-732 or their respective complements,  
comprising contacting a test agent with a colon or pancreatic cell and  
monitoring expression of the polynucleotide, wherein the test agent which  
modifies the expression of the polynucleotide is a candidate agent.

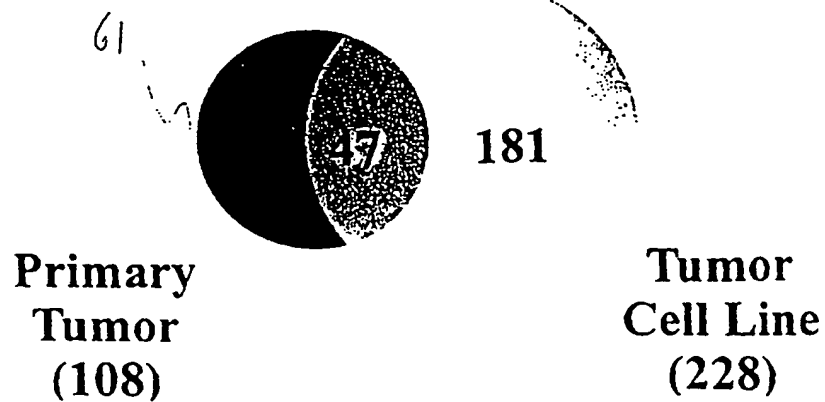
1/2



B.



C.

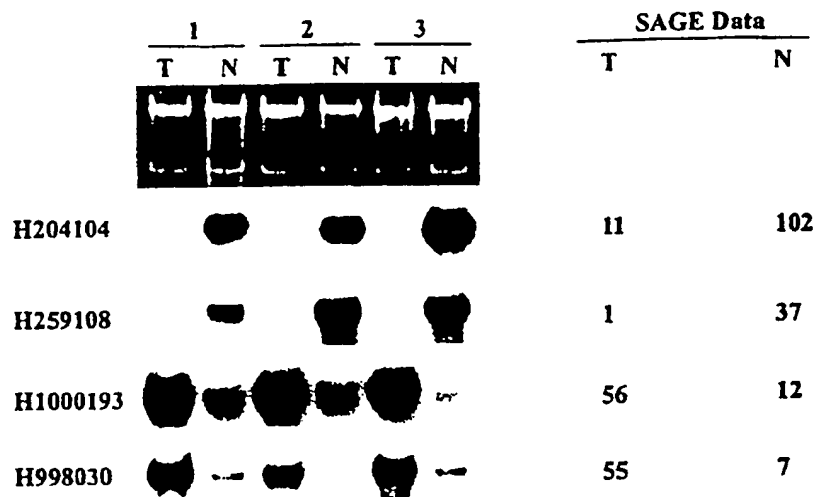




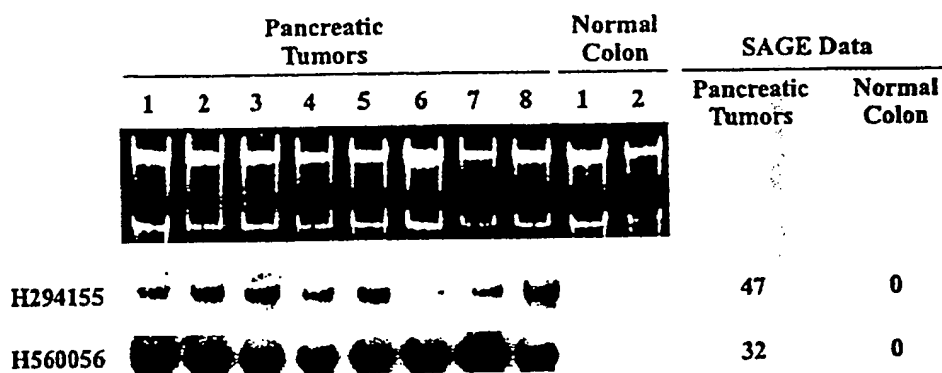
2/2

FIG. 2

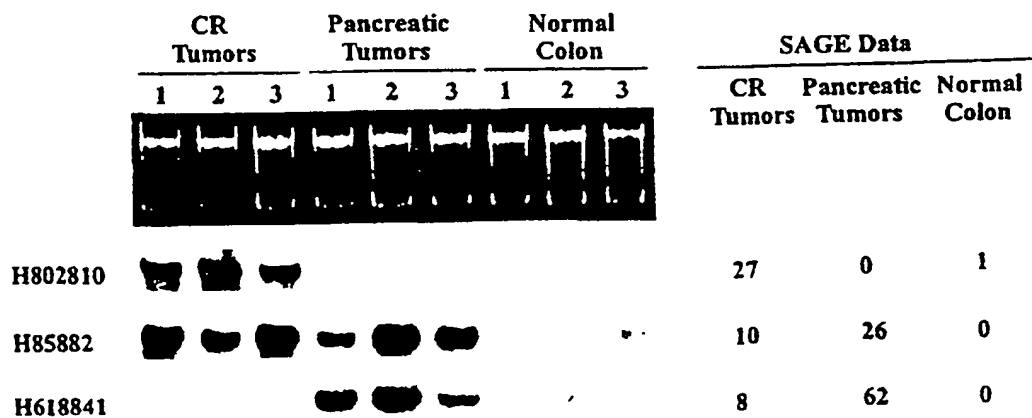
A.



B.



C.





## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup>:</b>  <b>C12Q 1/68, G01N 33/574</b>	<b>A3</b>	<b>(11) International Publication Number:</b> <b>WO 98/53319</b>  <b>(43) International Publication Date:</b> 26 November 1998 (26.11.98)
<b>(21) International Application Number:</b> PCT/US98/10277  <b>(22) International Filing Date:</b> 20 May 1998 (20.05.98)  <b>(30) Priority Data:</b> 60/047,352      21 May 1997 (21.05.97)      US  <b>(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application</b> US      60/047,352 (CON) Filed on      21 May 1997 (21.05.97)  <b>(71) Applicant (for all designated States except US):</b> THE JOHNS HOPKINS UNIVERSITY [US/US]; Suite 2-100, 2024 E. Monument Street, Baltimore, MD 21205 (US).  <b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> VOGELSTEIN, Bert [US/US]; The Johns Hopkins University, Suite 2-100, 2024 E. Monument Street, Baltimore, MD 21205 (US). KINZLER, Kenneth, W. [US/US]; The Johns Hopkins University, Suite 2-100, 2024 E. Monument Street, Baltimore, MD 21205 (US).	<b>(74) Agents:</b> KAGAN, Sarah, A. et al.; Banner & Witcoff, Ltd., 11th floor, 1001 G Street, N.W., Washington, DC 20001-4597 (US).  <b>(81) Designated States:</b> AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).  <b>Published</b> <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>  <b>(88) Date of publication of the international search report:</b> 8 July 1999 (08.07.99)	
<b>(54) Title:</b> GENE EXPRESSION PROFILES IN NORMAL AND CANCER CELLS  <b>(57) Abstract</b>  As a step towards understanding the complex differences between normal and cancer cells, gene expression patterns were examined in gastrointestinal tumors. More than 300,000 transcripts derived from at least 45,000 different genes were analyzed. Although extensive similarity was noted between the expression profiles, more than 500 transcripts that were expressed at significantly different levels in normal and neoplastic cells were identified. These data provide insights into the extent of expression differences underlying malignancy and reveal genes that are useful as diagnostic or prognostic markers.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## INTERNATIONAL SEARCH REPORT

Inter: ional Application No

PCT/US 98/10277

A. CLASSIFICATION OF SUBJECT MATTER  
 IPC 6 C12Q1/68 G01N33/574

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12Q G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SUGIO K ET AL.: "Differential expression of c-myc gene and c-fos gene in premalignant and malignant tissues." CANCER RESEARCH, vol. 48, no. 17, 1988, pages 4855-4861, XP002089885 see the whole document	1,3,13, 16,17,28
X	VAN BELZEN N ET AL.: "Detection of different gene expression in differentiating colon carcinoma cells by differential display" JOURNAL OF PATHOLOGY, vol. 178, no. Suppl., - 1996 page 2A XP002089886 see abstract	1,3,5,7, 9,11
Y	---	26,28,34
	---	

-/-

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

## \* Special categories of cited documents:

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

\*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

\*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

\*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

\*Z\* document member of the same patent family

Date of the actual completion of the international search

13 January 1999

Date of mailing of the international search report

24.05.1999

Name and mailing address of the ISA

European Patent Office, P. B. 5818 Patentlaan 2  
 NL - 2280 HV Rijswijk  
 Tel. (+31-70) 340-2111, 31 651 6001  
 Fax: (+31-70) 340-2115

Authorized officer

Knehr, M

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 98/10277

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 95 21944 A (SMITHKLINE BEECHAM CORP ;ROSENBERG MARTIN (US); DEBOUCK CHRISTINE) 17 August 1995 see the whole document ---	26,28,34
Y	EP 0 284 362 A (ICI PLC) 28 September 1988  see abstract see page 2, line 44 - line 51 see page 10, line 12 - line 15; claims 1,9; figure 2 ---	1,3,5,7, 9,11, 13-23, 26,28, 34,52
Y	EP 0 761 822 A (UNIV JOHNS HOPKINS MED) 12 March 1997  see the whole document ---	1,3,5,7, 9,11, 13-23, 26,28, 34,52
Y	WO 95 11923 A (DANA FARBER CANCER INST INC ;CHEN LAN BO (US); BAO SHIDENG (CN); L) 4 May 1995  see the whole document ---	1,3,5,7, 9,11, 13-18, 23,26, 28,34,52
Y	VELCULESCU V E ET AL: "SERIAL ANALYSIS OF GENE EXPRESSION" SCIENCE, vol. 270, 20 October 1995, pages 484-487, XP002053721 cited in the application see the whole document ---	1,3,5,7, 9,11, 13-18, 23,26, 28,34,52
Y	SCHWEINFEST C W ET AL.: "Subtraction hybridization cDNA libraries from colon carcinoma and hepatic cancer" GENETIC ANALYSIS TECHNIQUES AND APPLICATIONS, vol. 7, 1990, pages 64-70, XP002089887 see the whole document ---	1,3,5,7, 9,11, 13-18, 23,26
Y	WO 97 14812 A (CHIRON CORP) 24 April 1997 see the whole document ---	52
A	GRESS T M ET AL.: "A pancreatic cancer-specific expression profile" ONCOGENE, vol. 13, 1996, pages 1819-1830, XP002089888 see the whole document ---	
	--- -/--	

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 98/10277

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No
A	WO 95 19369 A (UNIV VANDERBILT) 20 July 1995 see the whole document ---	
A	GRESS T ET AL.: "Identification of genes with pancreatic cancer specific expression by use of cDNA representational difference analysis" GASTROENTEROLOGY, vol. 110, no. 4 Suppl., 1996, XP002089889 see abstract ---	
P,X	ZHANG L E AL.: "Gene expression profiles in normal and cancer cells." SCIENCE, vol. 276, 1997, pages 1268-1272, XP002089890 see the whole document ---	1,3,5,7, 9,11, 13-23, 26,28, 34,52
P,X	VAN BELZEN N ET AL.: "A novel gene which is up-regulated during colon epithelial cell differentiation and down-regulated in colorectal neoplasms" LABORATORY INVESTIGATION, vol. 77, no. 1, 1997, pages 85-92, XP002089891 see the whole document -----	1,3,5,7, 9,11,13, 14, 16-18, 23,26, 28,34

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 98/ 10277

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
  
2. ☐ Claims Nos.:  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
  
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see FURTHER INFORMATION sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
  
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
  
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
  
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

see FURTHER INFORMATION sheet, subject 1.

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

# INTERNATIONAL SEARCH REPORT

International Application No. PCT/ US 98/10277

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1,3,5,7,9,11,13-23,26,28,34,52 (partial)

**INVENTION 1:**

An isolated and purified human nucleic acid molecule comprising SEQ ID NO:291 of table 3 (INVENTION 1), an isolated nucleotide probe hybridizing to it, a diagnostic reagent comprising at least two of such probes, methods of diagnosing or prognosing colon cancer using them, and a method for screening for agents modulating the expression of such a sequence using them.

2. Claims: 1,3,5,7,9,11,13-23,26,28,34,52 (partial)

**INVENTION 2 to INVENTION 259:**

An isolated and purified human nucleic acid molecule comprising SEQ ID NO:292 of table 3 (INVENTION 2), an isolated nucleotide probe hybridizing to it, a diagnostic reagent comprising at least two of such probes, methods of diagnosing or prognosing colon cancer using them, and a method for screening for agents modulating the expression of such a sequence using them.

---

...ibidem for each of the SEQ ID Nos:293 to 549 (INVENTION 3 to INVENTION 259) as specified in table 3, separately.

3. Claims: 2,4,6,8,10,12-23,27,29,35,38-40,43,46,49, 52 (partial)

**INVENTION 260 to INVENTION 549:**

An isolated and purified human nucleic acid molecule comprising SEQ ID NO:1 of table 2 (INVENTION 260), an isolated nucleotide probe hybridizing to it, a diagnostic reagent comprising at least two of such probes, methods of diagnosing or prognosing colon cancer using them, a method of treating a cancer cell using them, an antibody linked to a cytotoxic agent used in such a method, and a method for screening for agents modulating the expression of such a sequence using them.

---

...ibidem for each of the SEQ ID Nos:2 to 290 (INVENTION 261 to INVENTION 549) as specified in table 2, separately.

4. Claims: 13-24,30,32,36,38,39,41,44,47,50,52 (partial)



# INTERNATIONAL SEARCH REPORT

International Application No. PCT/ US 98/10277

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

## INVENTION 550 to INVENTION 732:

An isolated and purified human nucleic acid molecule comprising SEQ ID NO:550 of table 4 (INVENTION 550), an isolated nucleotide probe hybridizing to it, a diagnostic reagent comprising at least two of such probes, methods of diagnosing or prognosing pancreatic cancer using them, a method of treating a cancer cell using them, an antibody linked to a cytotoxic agent used in such a method, and a method for screening for agents modulating the expression of such a sequence using them.

---

...ibidem for each of the SEQ ID Nos:551 to 732 (INVENTION 551 to INVENTION 732) as specified in table 4, separately.

5. Claims: 24,30,32,36,38,39,41,44,47,50 (partial)

## INVENTION 733 to INVENTION 734:

Methods of diagnosing or prognosing pancreatic cancer relying on a human nucleic acid molecule comprising SEQ ID NO:733 of table 4 (INVENTION 733), a method of treating a cancer cell using it, and an antibody linked to a cytotoxic agent used in such a method.

---

...ibidem for SEQ ID Nos:734 (INVENTION 734) as specified in table 4.

6. Claims: 25,31,33,37-39,42,45,48,51 (partial)

## INVENTION 735 to INVENTION 870:

Methods of diagnosing or prognosing cancer relying on a human nucleic acid molecule comprising SEQ ID NO:735 of table 5 (INVENTION 735), a method of treating a cancer cell using it, and an antibody linked to a cytotoxic agent used in such a method.

---

...ibidem for each of the SEQ ID Nos:736 to 870 (INVENTION 736 to INVENTION 870) as specified in table 5, separately.

## INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 98/10277

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9521944 A	17-08-1995	EP 0743989 A JP 9508800 T	27-11-1996 09-09-1997
EP 0284362 A	28-09-1988	AU 625169 B AU 1337888 A DK 159788 A FI 881388 A JP 1034291 A PT 87055 A,B	02-07-1992 22-09-1988 24-09-1988 24-09-1988 03-02-1989 01-04-1988
EP 0761822 A	12-03-1997	US 5695937 A US 5866330 A AU 6561496 A AU 7018896 A CA 2185379 A GB 2305241 A IE 80465 B JP 10511002 T WO 9710363 A	09-12-1997 02-02-1999 20-03-1997 01-04-1997 13-03-1997 02-04-1997 12-08-1998 27-10-1998 20-03-1999
WO 9511923 A	04-05-1995	CA 2175380 A EP 0725799 A US 5889159 A US 5872235 A	04-05-1995 14-08-1996 30-03-1999 16-02-1999
WO 9714812 A	24-04-1997	AU 7264196 A EP 0862651 A	07-05-1997 09-09-1998
WO 9519369 A	20-07-1995	US 5677125 A AU 1831795 A CA 2210396 A EP 0804453 A	14-10-1997 01-08-1995 20-07-1995 05-11-1997

Page 1

[illegible]

## Report Selection

- ☒ by Due Date  
☐ by Due Date (Calendar)  
☐ by Attorney  
☐ by Attorney (Via Email)  
☐ by Division

Report Format

- ☐ Concise  
☒ Comprehensive

Remarks

- ☐ Print  
☒ Do Not Print

Preview before printing: ☒From Date:  To Date: 30-Jun-2005Division:  Attorney: Action Type: Action Due: Annuity Due Local Agent Ref #: Indicator: Country:  Area: Status Codes: ☐ Active☐ Inactive☒ All

Status(es):

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449A/PTO

**INFORMATION DISCLOSURE  
STATEMENT BY APPLICANT**

(Use as many sheets as necessary)

Sheet 1 of 1

**Complete if Known**

Application Number	10/777,860
Filing Date	February 11, 2004
First Named Inventor	Chunmei Liu
Art Unit	1632
Examiner Name	Unknown
Attorney Docket Number	3581.1

**U.S. PATENT DOCUMENTS**

Examiner Initials *	Cite No. <sup>1</sup>	Document Number	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
		Number - Kind Code <sup>2</sup> (if known)			
	A1	US-6033860	3/7/2000	Lockhart et al.	
	A2	US- 6841348	1/11/2005	Warrington et al.	
	A3	US- 6884578	4/26/2005	Warrington et al.	
	A4	US- 2002142981A1	10/3/2002	Horne et al.	
	A5	US- 2003113762A1	6/19/2003	Warrington, Janet A.	
	A6	US- 2003175736A1	9/18/2003	Chinnaiyan et al.	
	A7	US- 2003180770A1	9/25/2003	Damokosh et al.	
	A8	US- 2003219771A1	11/27/2003	Bevilacqua et al.	
	A9	US- 2004009489A1	1/15/2004	Golub et al.	
	A10	US- 2004063120A1	4/1/2004	Beer et al.	
	A11	US- 2004126755A1	7/1/2004	Stephan et al.	
	A12	US- 2004126784A1	7/1/2004	Hitoshi et al.	
	A13	US- 2004146907A1	7/29/2004	Smith, Victoria	
	A14	US- 2004146921A1	7/29/2004	Eveleigh et al.	
	A15	US- 2004191819A1	9/30/2004	Eveleigh et al.	
	A16	US- 2004219565A1	11/4/2004	Kauppinen et al.	
	A17	US- 2004241696A1	12/2/2004	Peinado et al.	
	A18	US- 2005009067A1	1/13/2005	Logsdon et al.	
	A19	US- 2005032065A1	2/10/2005	Afar et al.	
	A20	US- 2005064472A1	3/24/2005	Shekar et al.	
	A21	US- 2005095607A1	5/5/2005	Erlander et al.	

**FOREIGN PATENT DOCUMENTS**

Examiner Initials *	Cite No. <sup>1</sup>	Foreign Patent Document	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear	T <sup>6</sup>
		Country Code <sup>3</sup> - Number <sup>4</sup> - Kind Code <sup>5</sup> (if known)				
	B1	EP1355149A2	10/22/2003	Wang, Yixin		
	B2	WO02101357A2	12/19/2002	SU et al.		
	B3	WO04033641A2	4/22/2004	SANCHEZ- CARBAYO et al.		
	B4	WO04066941A2	8/12/2004	Eveleigh et al.		
	B5	WO04081564A1	9/23/2004	BOWTELL et al.		
	B6	WO04105573A2	12/9/2004	SHOWE et al.		
	B7	WO05032495A2	4/14/2005	TAYLOR et al.		
	B8	WO9853319A2	11/26/1998	VOGELSTEIN et al.		

Examiner  
SignatureDate  
Considered

\*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. <sup>1</sup> Applicant's unique citation designation number (optional). <sup>2</sup> See Kinds Codes of USPTO Patent Documents at [www.uspto.gov](http://www.uspto.gov) or MPEP 901.04. <sup>3</sup> Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). <sup>4</sup> For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. <sup>5</sup> Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST. 16 if possible. <sup>6</sup> Applicant is to place a check mark here if English language Translation is attached.

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.



Page: 1

Due Date Indicator	Action Due Action Type	Case Number/SubCase Country	Status Division	App Number Pat Number	App Date Iss Date	Attys
12-Jul-2005	Annuity Due	3214.1 /	Published	00114152.2	12-Jul-2000	SW
Due Date	FOR-Annuity Due	European Patent Convention	Affymetrix			SK
Local Agent Ref #:	03848.00010	Title: Genotyping Biallelic Markers				
13-Jul-2005	Annuity Due	3214.1 /	Pending	2,313,869	13-Jul-2000	SW
Due Date	FOR-Annuity Due	Canada	Affymetrix			SK
Local Agent Ref #:	03848.0008	Title: Genotyping Biallelic Markers				
13-Jul-2005	Annuity Due	3429.1 /	Published	03764652.8	14-Jul-2003	TM
Due Date	FOR-Annuity Due	European Patent Convention	Affymetrix			LT
Local Agent Ref #:	AFMX-PEP-032	Title: System and Method for Examination of Microarrays Using Scanning Electron Microscope				
14-Jul-2005	Annuity Due	3429.1 /	Pending	2,492,281	14-Jul-2003	TM
Due Date	FOR-Annuity Due	Canada	Affymetrix			LT
Local Agent Ref #:	AFMX-PA-032	Title: System and Method for Examination of Microarrays Using Scanning Electron Microscope				
14-Jul-2005	Annuity Due	3507.1 /	Pending	2,492,212	14-Jul-2003	TM
Due Date	FOR-Annuity Due	Canada	Affymetrix			LT
Local Agent Ref #:	AFMX-PCA-025	Title: Nucleic Acid Labeling Methods				
14-Jul-2005	Annuity Due	3507.1 /	Published	03764653.6	14-Jul-2003	TM
Due Date	FOR-Annuity Due	European Patent Convention	Affymetrix			LT
Local Agent Ref #:	AFMX-PEP-025	Title: Nucleic Acid Labeling Methods				
16-Jul-2005	Annuity Due	1000.1A /	Granted	96-705897	21-Oct-1996	PLM
Due Date	FOR-Annuity Due	Korea, Republic of	Affymetrix	118362	16-Jul-1997	JL
Local Agent Ref #:	16528X-1-1KR	Title: Very Large Scale Immobilized Polymer Synthesis				
20-Jul-2005	Annuity Due	3014.2/PCT /	Pending	2,339,016	20-Jul-1999	TM
Due Date	FOR-Annuity Due	Canada	Affymetrix			WW
Local Agent Ref #:	56297-5035CA	Title: Nucleic Acid Labeling Compounds				
20-Jul-2005	Annuity Due	3014.2/PCT /	Granted	99937150.3	20-Jul-1999	TM
Due Date	FOR-Annuity Due	European Patent Convention	Affymetrix	1124838	01-Jun-2005	WW
Local Agent Ref #:	56297-5035EP	Title: Nucleic Acid Labeling Compounds				

Page 2

Due Date Indicator	Action Due Action Type	Case Number/SubCase Country	Status Division	App Number Pat Number	App Date Iss Date	Attys
22-Jul-2005 Due Date	Annuity Due FOR-Annuity Due	1000.1 / Korea, Republic of	Granted Affymetrix	91-701791 118659	07-Jun-1990 11-Feb-1997	PLM JL
Local Agent Ref #: 18547-000110KR Title: Large Scale Photolithographic Solid Phase Synthesis of Polypeptides and Receptor Binding Screening Thereof						
24-Jul-2005 Due Date	Annuity Due FOR-Annuity Due	3030.1/PCT / European Patent Convention	Pending Affymetrix	98936972.3	24-Jul-1998	WZ JL
Local Agent Ref #: 18547-035010EP Title: Gene Expression and Evaluation System						
24-Jul-2005 Due Date	Annuity Due FOR-Annuity Due	3068.1/PCT / Belgium	Granted Affymetrix	98936012.8 1002264	17-Feb-2000 14-Apr-2004	WZ JL
Local Agent Ref #: 018547-033810EP Title: Method and Apparatus for Providing a Bioinformatics Database						
24-Jul-2005 Due Date	Annuity Due FOR-Annuity Due	3068.1/PCT / Switzerland	Granted Affymetrix	98936012.8 1002264	17-Feb-2000 14-Apr-2004	WZ JL
Local Agent Ref #: 018547-033810EP Title: Method and Apparatus for Providing a Bioinformatics Database						
24-Jul-2005 Due Date	Annuity Due FOR-Annuity Due	3068.1/PCT / Germany	Granted Affymetrix	98936012.8 69823206.2	17-Feb-2000 14-Apr-2004	WZ JL
Local Agent Ref #: 018547-033810EP Title: Method and Apparatus for Providing a Bioinformatics Database						
24-Jul-2005 Due Date	Annuity Due FOR-Annuity Due	3068.1/PCT / European Patent Convention	Granted Affymetrix	98936012.8 1002264	17-Feb-2000 14-Apr-2004	WZ JL
Local Agent Ref #: 018547-033810EP Title: Method and Apparatus for Providing a Bioinformatics Database						
24-Jul-2005 Due Date	Annuity Due FOR-Annuity Due	3068.1/PCT / France	Granted Affymetrix	98936012.8 1002264	17-Feb-2000 14-Apr-2004	WZ JL
Local Agent Ref #: 018547-033810EP Title: Method and Apparatus for Providing a Bioinformatics Database						
24-Jul-2005 Due Date	Annuity Due FOR-Annuity Due	3068.1/PCT / United Kingdom	Granted Affymetrix	98936012.8 1002264	17-Feb-2000 14-Apr-2004	WZ JL
Local Agent Ref #: 018547-033810EP Title: Method and Apparatus for Providing a Bioinformatics Database						
24-Jul-2005 Due Date	Annuity Due FOR-Annuity Due	3068.1/PCT / Ireland	Granted Affymetrix	98936012.8 1002264	17-Feb-2000 14-Apr-2004	WZ JL
Local Agent Ref #: 018547-033810EP Title: Method and Apparatus for Providing a Bioinformatics Database						



**Page: 3**

<b>Due Date Indicator</b>	<b>Action Due Action Type</b>	<b>Case Number/SubCountry</b>	<b>Status Division</b>	<b>App Number Pat Number</b>	<b>App Date Iss Date</b>	<b>Attys</b>
<b>24-Jul-2005</b>	Annuity Due	3068.1/PCT / Italy	Granted Affymetrix	98936012.8	17-Feb-2000	WZ
<b>Local Agent Ref #: 018547-033810EP</b>		<b>Title:</b> Method and Apparatus for Providing a Bioinformatics Database				
<b>24-Jul-2005</b>	Annuity Due	3068.1/PCT / Netherlands	Granted Affymetrix	98936012.8	17-Feb-2000	WZ
<b>Local Agent Ref #: 018547-033810EP</b>		<b>Title:</b> Method and Apparatus for Providing a Bioinformatics Database				
<b>24-Jul-2005</b>	Annuity Due	3068.1A / European Patent Convention	PUBLISHED AFFYMETRIX	03025794.3	24-Jul-1998	WZ  JL
<b>Local Agent Ref #: 18547-033811EP</b>		<b>Title:</b> Method and Apparatus for Providing a Bioinformatics Database				
<b>24-Jul-2005</b>	Annuity Due	3068.2/PCT / European Patent Convention	PENDING AFFYMETRIX	98937128.1	18-Feb-2000	WZ  JL
<b>Local Agent Ref #: 18547-033820/EP</b>		<b>Title:</b> Method and System for Providing a Polymorphism Database				
<b>24-Jul-2005</b>	Annuity Due	3068.3/PCT / European Patent Convention	PENDING AFFYMETRIX	98937126.5	24-Jul-1998	WZ  JL
<b>Local Agent Ref #: 18547-033830EP</b>		<b>Title:</b> Method and System for Providing a Probe Array Chip Design Database				

# Report Selection

- ☒ by Due Date
- ☐ by Due Date (Calendar)
- ☐ by Attorney
- ☐ by Attorney (Via Email)
- ☐ by Division

Report Format

- ☐ Concise
- ☒ Comprehensive

Remarks

- ☐ Print
- ☒ Do Not Print

Preview before printing: ☒

From Date: 30-Jun-2005 To Date: 31-Jul-2005

Division: Attorney:

Action Type:

Action Due: Annuity Due

Local Agent Ref #:

Indicator:

Country: Area:

Status Codes:

- ☐ Active
- ☐ Inactive
- ☒ All

Status(es):

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☒ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**